

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:45:08 ; Search time 25.29 Seconds  
(without alignments)  
890.486 Million cell updates/sec

Title: US-09-762-926-4  
Perfect score: 4904  
Sequence: 1 MRSSFLKPKICFYLMGVMLY.....SVLTNFARGRFLTIMSKYF 922

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues

Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/1aa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/1aa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/1aa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/1aa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/1aa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/1aa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	284	5.8	915	3 US-08-613-009A-18	Sequence 18, Appl
2	278	5.7	915	1 US-08-487-890A-96	Sequence 96, Appl
3	278	5.7	915	2 US-08-363-124A-2	Sequence 2, Appli
4	278	5.7	915	2 US-08-478-435-96	Sequence 96, Appl
5	278	5.7	915	2 US-08-337-483-96	Sequence 96, Appl
6	278	5.7	915	2 US-08-478-373-96	Sequence 96, Appl
7	278	5.7	915	3 US-08-474-671-96	Sequence 96, Appl
8	278	5.7	915	3 US-08-483-577A-96	Sequence 96, Appl
9	278	5.7	915	4 US-08-897-438-96	Sequence 96, Appl
10	278	5.7	915	4 US-08-637-654-96	Sequence 96, Appl
11	269.5	5.5	908	1 US-08-487-890A-94	Sequence 94, Appl
12	269.5	5.5	908	2 US-08-478-435-94	Sequence 94, Appl
13	269.5	5.5	908	2 US-08-337-483-94	Sequence 94, Appl
14	269.5	5.5	908	2 US-08-478-373-94	Sequence 94, Appl
15	269.5	5.5	908	3 US-08-474-671-94	Sequence 94, Appl
16	269.5	5.5	908	3 US-08-483-577A-94	Sequence 94, Appl
17	269.5	5.5	908	3 US-08-448-194-4	Sequence 4, Appli
18	269.5	5.5	908	3 US-08-613-009A-16	Sequence 16, Appl
19	269.5	5.5	908	4 US-08-897-438-94	Sequence 94, Appl
20	269.5	5.5	908	4 US-08-867-921-4	Sequence 4, Appli
21	269.5	5.5	908	4 US-08-637-654-94	Sequence 94, Appl
22	261	5.3	909	2 US-08-363-124A-4	Sequence 4, Appli
23	260	5.3	791	3 US-08-537-361E-4	Sequence 4, Appli
24	257	5.2	791	4 US-08-817-707-4	Sequence 4, Appli
25	257	5.2	911	1 US-08-487-890A-95	Sequence 95, Appl
26	257	5.2	911	2 US-08-478-435-95	Sequence 95, Appl
27	257	5.2	911	2 US-08-337-483-95	Sequence 95, Appl

28	257	5.2	911	2	US-08-478-373-95	Sequence 95, Appl
29	257	5.2	911	3	US-08-474-671-95	Sequence 95, Appl
30	257	5.2	911	3	US-08-483-577A-95	Sequence 95, Appl
31	257	5.2	911	3	US-08-448-194-6	Sequence 6, Appli
32	257	5.2	911	3	US-08-613-009A-17	Sequence 17, Appl
33	257	5.2	911	4	US-08-897-438-95	Sequence 95, Appl
34	257	5.2	911	4	US-08-867-921-6	Sequence 6, Appli
35	257	5.2	911	4	US-08-637-654-95	Sequence 95, Appl
36	255	5.2	790	4	US-08-817-707-6	Sequence 6, Appli
37	250	5.1	884	1	US-08-066-167-2	Sequence 2, Appli
38	243.5	5.0	941	4	US-09-074-658-75	Sequence 75, Appl
39	242	4.9	790	4	US-08-817-707-8	Sequence 8, Appli
40	239	4.9	790	3	US-08-537-361E-8	Sequence 15, Appl
41	239	4.9	792	3	US-08-613-009A-15	Sequence 6, Appli
42	238.5	4.9	792	3	US-08-537-361E-6	Sequence 6, Appli
43	237.5	4.8	792	1	US-08-326-670A-2	Sequence 2, Appli
44	237.5	4.8	792	3	US-08-990-470A-2	Sequence 2, Appli
45	237.5	4.8	792	4	US-08-817-707-2	Sequence 2, Appli

ALIGNMENTS

RESULT 1  
US-08-613-009A-18  
; Sequence 18, Application US/08613009A  
; Patent No. 6090576  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM: disk  
; MEDIUM TYPE: Floppy disk  
; OPERATING SYSTEM: IBM PC compatible  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; FILING DATE: 08-MAR-1996  
; APPLICATION NUMBER: US/08/613,009A  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-542  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 915 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-613-009A-18

Query Match 5.8%; Score 284; DB 3; Length 915;  
Best Local Similarity 21.5%; Pred. No. 1.3e-15;  
Matches 225; Conservative 128; Mismatches 321; Indels 374; Gaps 59;

QY 5 FRLKPICFYLMGVNLYHHSYAEDAGRAGSEAOIQVLEDDVHVAK--RVPKDKKVFVDARA 62  
Db 7 FRLNLCISLMTAL---PAYAENV-QAG-QAQBKQDITQVKKAKKQTRDRNEVTGLGKL 61  
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFRVNTMVDGIT 118  
Db 62 VKTADTSLKEQVLDIRDLTRYDPCIAVVEOGRGASSG---YSIRG-MDKNRVSLTVDGLA 117  
QY 119 Q-TFYSTSTDA--RAGSSQFASVDSNFTAGLDVVVKGFSGAGINSAGSANTLTLG 175  
Db 118 QIQSYTAQAALGGTRTAGSGGAINIEYENVKAVEISKGSNSVEQSGGALAGSVAFOTKT 177  
QY 176 VDDVVOGNNTYGLLLK-GLTGTSN--TKGNAMAA-IGARKWL--ESGASVGLYGHRSRS 229  
Db 178 ADDVIGEGRWGQIQSKTAYSGKNRGLTQSLALAGRIGGAEALLIRTGRAHRAH---- 233  
QY 230 VAQYRVGGGQHGHNFGAEYLERKORYFVQEGGLFNFSNKGWRDQFQYKTKWYQ 289  
Db 234 -----EAAGRGVQSFNPLAPVDDGSKYAYFIVEECK-----NGHEK-----C 272  
QY 290 KYNDPQELQYIEGHDK---SWRENLAPOVDITPIDPSSILKOOS-----AGNLF--KLEY 339  
Db 273 KANPPKD---VVGEDKQVSTREDYTPNRFLEA--DPLSVESRSLFRPGFREKNRHY 326  
QY 340 DGVFNKYTAQ---PRDLNTKIGSKII-----NRNVPNYGLSLNSYANLNL 383  
Db 327 IGGILERTQOTFTDRMTVPAFLTKAVFDANQKQAGSLRGKYGAGNH-----KYGGL-- 379  
QY 384 TAAVNSGRQKYPKSGKFTGGLLKDFETYNNAKILDLNNTATPLPRETELOITLGNFYF 443  
Db 380 ---FTSGENNAVGAEY-GGVFVD-----ETHTKSYGLEVY 413  
QY 444 HNEYGKNRFPPEELGFED--GPDDNGL-----YSYLCRFRKGDK 480  
Db 414 YTNADKDTWADYARLSYDRQIGLDNFQTHCSADGSKYCRPSADKPSY---YKSDR 470  
QY 481 GLLPKSTIVQAGSQVNTFYDPAALKKDIYRLNYSNTNVGY-RFCG-----EYTYGG 534  
Db 471 VIYGESHKLLQAAPKKSFDATAI-----RHNLSVN-LGYDRFGSLNRHQDY---YYQ 518  
QY 535 SDDEFKRAFGENSPYKKHCKNOSGCIYEPVLKYGKKRA-----NHNYSVISAD-- 583  
Db 519 SAN---RAYSKTTP-----QNGKATSPNGRKKNPYVWSIGRGNVVR 559  
QY 584 ---FGD-----YFMPPA-----SYSRTH----- 598  
Db 560 QICLFGNNTVDCPRINSKSYAAVRDNRVLGRWADVAGLRYDYRSTHSDDGSVSTG 619  
QY 599 -----RMPNIQEMFYSOIGDSGVHTALKPERANTWQ 629  
Db 620 THRTLWSNAGIVLKPADWLDTYRTSTGFRPLPSFAEMYGWRSGDKIKAVKIDPEKSFNE 679  
QY 630 FG-----FNTYKKGILLKODDITLGLKLVGYRSRIDN-----YIH--- 662  
Db 680 AGIVFKGDFGNLEASWFNNAYRDLVR-----GYEAOIKDGKEQVKGNGPAYLNAQS 730  
QY 663 -----NVYGG--W---WDLNGNIP-SWVSSTGLAYT-IQHRNFKDKVHKHGFELNLD 709  
Db 731 ARITGINILCKIDWNGYWD--KLPEGWYST--PAYNRVRVRDIKKRADRTDQSHL-FD 784  
QY 710 YGRFTNLVAYOKSTOPTNFSASESPNASKEDQLKQYGLSRVSALPRDYGRLEVGT 769  
Db 785 -----AIQPSRYV-----VGSYDQPEGKGVNGMLT-----YSKAKEIT 819  
QY 770 RWLGNKLTGGMARYFCKSIATAEERYIDGTNGGNTSNVRLQKRSIKOTETLARQPLI 829  
Db 820 ELGSRALLNG-----NSRNTKATSRRTRPWYI 847  
QY 830 FDEYAAEYKPKNLIFRAEVENKLFDRRYI 857  
Db 848 VDVSGYTYVKKHFTLRAGVYNLLNHRIV 875

RESULT 2  
US-08-487-890A-96  
; Sequence 96, Application US/08487890A  
; Patent No. 5708149  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena  
; APPLICANT: Harkness, Robin  
; APPLICANT: Schryvers, Anthony  
; APPLICANT: Chong, Pele  
; APPLICANT: Gray-Owen, Scott  
; APPLICANT: Yang, Yan-ping  
; APPLICANT: Murdin, Andrew  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Transferrin Receptor Genes  
; NUMBER OF SEQUENCES: 147  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patent In Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/487,890A  
; FILING DATE: 07-JUN-1993  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/175,116  
; FILING DATE: 29-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/148,968  
; FILING DATE: 08-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 96:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 915 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-487-890A-96

Query Match 5.7%; Score 278; DB 1; Length 915;  
Best Local Similarity 21.4%; Pred. No. 4,4e-15;  
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;  
QY 5 FRLKPICFYLMGVNLYHHSYAEDAGRAGSEAOIQVLEDDVHVAK--RVPKDKKVFVDARA 62  
Db 7 FRLNLCISLMTAL---PAYAENV-QAG-QAQBKQDITQVKKAKKQTRDRNEVTGLGKL 61  
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFRVNTMVDGIT 118  
Db 62 VKTADTSLKEQVLDIRDLTRYDPCIAVVEOGRGASSG---YSIRG-MDKNRVSLTVDGLA 117  
QY 119 Q-TFYSTSTDA--RAGSSQFASVDSNFTAGLDVVVKGFSGAGINSAGSANTLTLG 175  
Db 118 QIQSYTAQAALGGTRTAGSGGAINIEYENVKAVEISKGSNSVEQSGGALAGSVAFOTKT 177  
QY 176 VDDVVOGNNTYGLLLK-GLTGTSN--TKGNAMAA-IGARKWL--ESGASVGLYGHRSRS 229  
Db 178 ADDVIGEGRWGQIQSKTAYSGKNRGLTQSLALAGRIGGAEALLIRTGRAHRAH---- 233

```
QY 230 VAQNYRVGGGQHGIGNFGAEYLERRKORYFVQEGGLKFNSNGKWERDORFPYKTKWYQ 289
Db 234 -----EAGRGVGSFNLAPVDGSKYAYFIVEECK-----NGGHECKKAMP-----276
QY 290 KYNDDPQELQYIEGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AGNLF--KLEY 339
Db 277 -----KDDVGEDKQVSTRDYTPNRF--DPLSYESRSWLFRCGFRFENKRYH 326
QY 340 DGVFNKYTAQ---FRDLNTKIGSRKII-----NRNYQFNYGLSLSYANLNL 383
Db 327 IGGILERTQOTFTDTRDMTPAFLTKAQVFNKQAGSLRGNGKYAGNH-----KYGGL--379
QY 384 TAAYNSGRQYKPGSKFTGCLLKDFETYNNAKILDLNNTATFRLPRETELQTTLGNFY 443
Db 380 ---FTSGENNAVGAAY-GTGVEYD-----ETHTKSRYGLEVY 413
QY 444 HNEYGKNRFPPEELGLFFD--GPODNGL-----YSYLGRPKGDK 480
Db 414 YTNADKDTWADYARLSYDRQIGLDNHFQTHCSADGSKYCRPSADKPSY---YKSDR 470
QY 481 GLLPQKSTIVOPAGSQVNTFYFDAAKKDIYRLNYSNTVGY-RFGG-----EYTGYYG 534
Db 471 VIYGESHKLLQAAFKKSFDTAKI-----RHNLNVN-LGYDRFGSGLNRHODY--YYQ 518
QY 535 SDDEFKAFGENSPYKKHCKNQSGIYEPVLKKGKRA-----NNHVSISAD-----583
Db 519 SAN---RAYSLKTPP-----QNGKKTSPNGREKNPYWISIGRGNVYTR 559
QY 584 ---FGD-----YFMPPA-----SYSRTH-----598
Db 560 QICLFGNNTYDCTPSRINSKSYAAVRDNRVLRGWADVGAGLYRSTHSDDGSVSTG 619
QY 599 -----RMPNIQEMYFSQIGDSGVHTALKPERANTWQ 629
Db 620 THRTLSNAGIVLKPADWLDLTYRSTGFRPLSPFAEMYGWRSGDKIKAVKIDPEKSENKE 679
QY 630 FG-----PNTYKGLKQDDTLGLKLYCYRSRIDN-----YIH---662
Db 680 AGIVEKGDGNLNASWFNNAYRDLIV-----GYEAQIKDGKEQVKGNGPAYLNAQS 730
QY 663 -----NVYKG--W---WDLGNIP-SWVSSTGLAYT-IQHRNFKDKVHKHGFLELNYD 709
Db 731 ARITGINILKIDWGVMD--KLPEGWYST--PAYNRVVRDIKKRADRTDIQSHL-FD 784
QY 710 YGRFTNLSTAYOKSTOPTWFSASESPNNASKEDQLKQYGLSRVSALPRDYGRLEVGT 769
Db 785 -----AIQPSRYV-----VCSGYDQPEGKWGVNGMLT---YSKAKEIT 819
QY 770 RWLGNKLTGAMRYFGKSIIRATAEERYIDGTNGGNTSNVRLQKRSIKQTEILARQPLI 829
Db 820 ELGSRALLNG-----NSRNTKATARTRPWYI 847
QY 830 FDFYAAEYPPKKNLIFRAEVKNLFDRRYI 857
Db 848 VDVSYYTVKKHFTLRAGVYNLLNHRVY 875

RESULT 3
US-08-363-124A-2
; Sequence 2, Application US/08363124A
; Patent No. 5912336
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding Proteins From
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Imclone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
```

```
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA: US/08/363,124A
; APPLICATION NUMBER: US/08/363,124A
; FILING DATE: 23-DEC-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/124,254
; FILING DATE: 20-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-363-124A-2
```

Query Match 5.7% Score 278; DB 2; Length 915;

Best Local Similarity 21.4%; Pred. No. 4.4e-15; Indels 374; Gaps 58;

Matches 224; Conservative 127; Mismatches 323;

```
QY 5 FRLKPICFYLMGVNLYHHSYAEDAGRAGEAQIQVLEDVHVYKAK--RVPDKKVFETDARA 62
Db 7 FRNLILSLMTAL---PAYAENV-QAG-QAQEKQOLDTIOVKAQKQKTRDRNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFRVNTWVDGIT 118
Db 62 VKTADTLSEQVLDIRDTRYDPIAVVEQGRGASSG---YSIRG-MDKNRVSTVVDGLA 117
QY 119 Q-TFYSTSDAG--RAGSSQFGASVDSNFTAGLDVVKGSFGSAGINSLAGSANRLTG 175
Db 118 QIOSYTAQAALGGTRTAGSSGAINETIENYKAVEISKGSNSVEQSGGALAGSAVFOTKT 177
QY 176 VDDVVOGNNTYGLLLK-GLTGTNS--TKGNAMAA-IGARKWL--ESGASGVLYGHRSRS 229
Db 178 ADVVIGRGQWGIQSKTAYSGKNRGLTQSIALAGRIGGAELLRTGRHAGEIRAH----233
QY 230 VAQNYRVGGGQHGIGNFGAEYLERRKORYFVQEGGLKFNSNGKWERDORFPYKTKWYQ 289
Db 234 -----EAGRGVGSFNLAPVDGSKYAYFIVEECK-----NGGHECKKAMP-----276
QY 290 KYNDDPQELQYIEGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AGNLF--KLEY 339
Db 277 -----KDDVGEDKQVSTRDYTPNRF--DPLSYESRSWLFRCGFRFENKRYH 326
QY 340 DGVFNKYTAQ---FRDLNTKIGSRKII-----NRNYQFNYGLSLSYANLNL 383
Db 327 IGGILERTQOTFTDTRDMTPAFLTKAQVFNKQAGSLRGNGKYAGNH-----KYGGL--379
QY 384 TAAYNSGRQYKPGSKFTGCLLKDFETYNNAKILDLNNTATFRLPRETELQTTLGNFY 443
Db 380 ---FTSGENNAVGAAY-GTGVEYD-----ETHTKSRYGLEVY 413
QY 444 HNEYGKNRFPPEELGLFFD--GPODNGL-----YSYLGRPKGDK 480
```

Db 414 YTNADKDTWADYARLSYDROGIGLDNHFQOHCSDAGSDKYCRPSADKPFY---YKSDR 470  
QY 481 GLLPQKSTIVOPAGSQVFNFTFYDAALKDYLRLNYSNTVGY-RFGG-----EYTGYYG 534  
Db 471 VIYGESHKLQAAPKKSFTAKI-----RHNLNVN-LGYDRFGSNLRHODY--YYQ 518  
QY 535 SDDEFKRAFGENSPYKHKCNQSCGIYEPVLKYYGKKRA-----NNHVSYSISAD----- 583  
Db 519 SAN---RAYSLKTPP-----QNNKKTSPPNGREKNPYWVSIGRGNVVT 559  
QY 584 ---FGD-----YFMPFA-----SYSRTH----- 598  
Db 560 QICLFGNNYTDCTPRINSKSYAAAVRDNVRLGRWADVAGLRYDYRSTHSDGVSYG 619  
QY 599 -----RMPNIQEMFYSGISDGVHTALKPERANTWQ 629  
Db 620 THRTLSNAGIVLKPADWLDITYRTSTGFRLPSPAEMYGWRSGDKIKAVKIDPEKSFENKE 679  
QY 630 FG-----FNTYKGLLKODDTLGLKLVGYRSDN-----YIH--- 662  
Db 680 AGIVFKGDFGNLEASWFNAYRDLIVR-----GYEAQIKDGKEQVKGNPAYLNAQS 730  
QY 663 ---NVYGG--W---WDLNGNIP-SWVSSTGLAYT-IOHRNFKDKVKHKGFELELNYD 709  
Db 731 ARITGINILKIDWNGVD---KLPEGWYST--PAYNRVRDIKKRADRTDIQSHL-FD 784  
QY 710 YGREFTNLSYAYOKSTQPTNFSDASEPNNAKEDQLKOGYGLSRVSALPRDYGRLEVGT 769  
Db 785 -----AIQPSRYV-----VGSYDQPEGWGVNGMLT-----YSKAKEIT 819  
QY 770 RWLGNKLTLCGAMRYFKGSRATRAERYIDCTNGGNTSNVRQLGKRSIKOTETILARQPLI 829  
Db 820 ELLGSRLANG-----NSRNTKATARTREPWI 847  
QY 830 DFVAAVEPKNLIFFRAEVKNLFDPRYI 857  
Db 848 VDVSQYTVVKKHFTLRAGVYNLLNHRIV 875

RESULT 4  
US-08-478-435-96  
; Sequence 96, Application US/08478435  
; Patent No. 5922323  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena  
; APPLICANT: Harkness, Robin  
; APPLICANT: Schryvers, Anthony  
; APPLICANT: Chong, Pele  
; APPLICANT: Gray-Owen, Scott  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Murdin, Andrew  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Transferrin Receptor Genes  
; NUMBER OF SEQUENCES: 147  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,435  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/337,483  
; FILING DATE: 08-NOV-1994

; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/175,116  
; FILING DATE: 29-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/148,968  
; FILING DATE: 08-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-462 MIS:Vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 96:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 915 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-478-435-96

Query Match 5.7%; Score 278; DB 2; Length 915;  
Best Local Similarity 21.4%; Pred. No. 4.4e-15;  
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

QY 5 FRKPCIFCLMGVLMVHSHYAEADAGAGSEAOIQVLEDVHVAK--RVPKDKKVFETDARA 62  
Db 7 FRNLICLSLMTAL---PAYAENV-QAG-QAQEKQLDTIOVKKAKKQKTRDNEVTGLGKL 61  
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQDQK--SSGIVSLNIRGDSGFGORVNTMVDGIT 118  
Db 62 VKTADTLSKEQVLDIRDLTRYPGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVGLA 117  
QY 119 Q-TFYSTSTTAG--RAGGSSQFGASVDSNFIAGLDVVVKGSFSSAGINSLAGSANLRTLG 175  
Db 118 QIOSYTAQAALGCTRTAGSSGAINETEVENVKAVEISKSGNSVEQSGALAGSVAFOTKT 177  
QY 176 VDDVOGNNNTYGLLLK-GLTGNTS--TKGNAMAA--TGARKWL--ESGASGVLYGHSSRS 229  
Db 178 ADVVIGEGROWIQSTAYSGKNRGLTQSTALAGRIGGAELLIRGRHAGEIRAH---- 233  
QY 230 VAQNYRVGGGQHGIFNGAELERRKQRYFVQEGGLKFNNSGKWERDFORPYWTKWKYQ 289  
Db 234 -----EAGRGVQSFNRLAPVDDGSKYAYFIVEECK-----NGGHEKCKANP----- 276  
QY 290 KYNDPQELQYIEGHDK-----SHRENLAPOYDITPDPSSLKQOS-----AGNLF--KLEY 339  
Db 277 -----KKDVVGEDKROTVDYTPGNRFLA--DPLSYESRSMFLRFGFRFENKRHY 326  
QY 340 DGVFNKYTAQ---FRDLNTRKIGSRKII-----NRNYQFNYSGLNSYANLNL 383  
Db 327 IGGILERTQQTFTDRMTVPALTKAVFDANQQAQSLRNGKYGAGNH-----KYGGL-- 379  
QY 384 TAAYNNGRQYKPGSKFTGMLLKDFETYNNAKILDNLNTATFRLPRETELQTLTGNFY 443  
Db 380 ---FTSGENNAPVGAAY-GTGVEYD-----ETHTKSRYGLEYV 413  
QY 444 HNEYGNRPPEELGLFFD--GPQDNGL-----YSYLGREFGDK 480  
Db 414 YTNADKDTWADYARLSYDROGIGLDNHFQOHCSDAGSDKYCRPSADKPFY---YKSDR 470  
QY 481 GLLPQKSTIVOPAGSQVFNFTFYDAALKDYLRLNYSNTVGY-RFGG-----EYTGYYG 534  
Db 471 VIYGESHKLQAAPKKSFTAKI-----RHNLNVN-LGYDRFGSNLRHODY--YYQ 518  
QY 535 SDDEFKRAFGENSPYKHKCNQSCGIYEPVLKYYGKKRA-----NNHVSYSISAD----- 583  
Db 519 SAN---RAYSLKTPP-----QNNKKTSPPNGREKNPYWVSIGRGNVVT 559  
QY 584 ---FGD-----YFMPFA-----SYSRTH----- 598



Db 560 QICLFGNNTYDCTPRSRNGSKSYAAVRDNRVLRGWADVAGAGLRYDYRSTHSDGVSSTG 619  
QY 599 -----RMPNIOEMFYSGIGDSGVHTALKPERANTWQ 629  
Db 620 THRTLSNAGIVLKPADWLDLTYRTSTGFRLPFAEMYGWRSGDKIKAVKIDPEKSNKE 679  
QY 630 FG-----FNTYKGLLKQDDTLGLKLVGYRSRDN-----YIH--- 662  
Db 680 AGIVFKGDFGNLEASFNNAYRDLIVR-----GYEAQIKDGKEQVKNPAYLNAQS 730  
QY 663 -----NVYCK--W---WDLNGNIP--SWVSSTGLAYT--IOHRNFKDKVHKHGFLELNYD 709  
Db 731 ARITGINILKIDWNGVD--KLPEGWYST--FAYNRVRVDIKKRADRTDIQSHL-FD 784  
QY 710 YGRFETNLVAYOKSTOPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGT 769  
Db 785 -----AIQPSRYV-----VSGGYDQEGKGVNGMLT-----YSKAKEIT 819  
QY 770 RWLGKNTLGGAMRYFGKSRATAERYIDGTNGGNTSNVRQLGKRSIKOTETLAROPLI 829  
Db 820 ELLGSRALNG-----NSRNTKATARTRPWYI 847  
QY 830 FDFYAAYPEKKNLIFRAEVKNLFDPRVYI 857  
Db 848 VDVSYYTVKKHETLRAGVYNLLNHRIV 875  
  
RESULT 5  
US-08-337-483-96  
; Sequence 96, Application US/08337483  
; Patent No. 592562  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena  
; APPLICANT: Harkness, Robin  
; APPLICANT: Schryvers, Anthony  
; APPLICANT: Chong, Pele  
; APPLICANT: Gray-Owen, Scott  
; APPLICANT: Yang, Yan-ping  
; APPLICANT: Murdin, Andrew  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Transferrin Receptor Genes  
; NUMBER OF SEQUENCES: 147  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 Unviersity Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/337,483  
; FILING DATE: 08-NOV-1994  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-410 MIS:1b  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 96:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 915 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
US-08-337-483-96

Query Match 5.7%; Score 278; DB 2; Length 915;  
Best Local Similarity 21.4%; Pred. No. 4.4e-15;  
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;  
  
QY 5 FRLPICFYLGMVMLYHHSYAEDAGRASEAQIOVLDDVHVAK--RVPDKKVFVTARA 62  
Db 7 FRNLILSLMTAL---PAYAENV-QAG-QAERKQDITQVAKKQKTRRDNEYTGLGKL 61  
QY 63 VSTRDIFKSS--NLDNIVRSIP--AFTQODK--SSGIVSLNIRGDSGFGVRNMTVDGIT 118  
Db 62 VKTADTLSEKQVLDIRDLTRYDPCIAVVEQGRGASS---YSIRG-WDKNRVSLTVDGLA 117  
QY 119 Q-TFYSTSTDAQ--RAGGSQFSGASVDSNFTAGLDVVVKGSGFSAGINSLASLANLTGL 175  
Db 118 QIQSYTAQAALGGTRTAGSSGAINIEYENVKAVEISKGSNSVQEGSAGALAGSVAFOTKT 177  
QY 176 VDDVQGNNTYGLLLK--CLTGTNS--TKGNMAA--IGARKWL--ESGASVCLVGHSSRS 229  
Db 178 ADDVIGEGROWGIQSKTAYSGKNRGLTQSIAGRIAGAEALLIRTGRAHEIRAH----- 233  
QY 230 VAQNYRVGGGQHIGNFGAEYLERRKQRYFVQEGGLFNNSGKWERDFQRPYKTKWYQ 289  
Db 234 -----EAAGRGVQSFNRLAPVDDGSKYAYFIVEECK-----NGGHECKANP----- 276  
QY 290 KYNDPQBLQYIEGHDK-----SWRENLAQYDITPIDPSSLKQOS-----AGNLF--KLEY 339  
Db 277 -----KKDVVGEDKQTVSTRDYTPNRFLA--DPLSYESRSWLFPRCFREKRRHY 326  
QY 340 DGVPNKYTAQ---FRDLNTKIGSRKII-----NRYQFNYGLSLNYSANLNL 383  
Db 327 IGGILERTQOTFTRDMTVPAFLTKAVFDANQKAGSLRGNGKYAGNH-----KYGLJ-- 379  
QY 384 TAAYNSGRQYKPGSKFTGMLKDFETYNNNAKILDLNNTATFRLPRETELQTTLGFNYF 443  
Db 380 ---FTSGENNAPVGAAY--GTGVFYD-----ETHKSRYLEYV 413  
QY 444 HNEYGKRRFPEELGLFED--GPDODNGL-----YSYLGRFKGDK 480  
Db 414 YTNADKDTWADYARLSYDRQIGLDNHFQOOTHCSADGSKYCRPSADKPFYS---YKSDR 470  
QY 481 GLLPQKSTIVPAGSQYFNTFYDAALKKDIYRLNYSTNTVGY--RFGG-----EYTVYG 534  
Db 471 VIYGESHKLQAQFKKSFDTAKI-----RHNLVN--LGYDRFGSNLRHODY--YYQ 518  
QY 535 SDDEFKRAFGENSPYKKHCNQSCGIYEPVLKYGKRA-----NNHSVSISAD----- 583  
Db 519 SAN---RAYSLKTPP-----QNNGKKTSPNGREKNPYWVSIGRGNVVTR 559  
QY 584 ---FGD-----YFMPEA-----SYSRTH----- 598  
Db 560 QICLFGNNTYDCTPRSRNGSKSYAAVRDNRVLRGWADVAGAGLRYDYRSTHSDGVSSTG 619  
QY 599 -----RMPNIOEMFYSGIGDSGVHTALKPERANTWQ 629  
Db 620 THRTLSNAGIVLKPADWLDLTYRTSTGFRLPFAEMYGWRSGDKIKAVKIDPEKSNKE 679  
QY 630 FG-----FNTYKGLLKQDDTLGLKLVGYRSRDN-----YIH--- 662  
Db 680 AGIVFKGDFGNLEASFNNAYRDLIVR-----GYEAQIKDGKEQVKNPAYLNAQS 730  
QY 663 -----NVYCK--W---WDLNGNIP--SWVSSTGLAYT--IOHRNFKDKVHKHGFLELNYD 709  
Db 731 ARITGINILKIDWNGVD--KLPEGWYST--FAYNRVRVDIKKRADRTDIQSHL-FD 784  
QY 710 YGRFETNLVAYOKSTOPTNFSDAESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGT 769  
Db 785 -----AIQPSRYV-----VSGGYDQEGKGVNGMLT-----YSKAKEIT 819  
QY 770 RWLGKNTLGGAMRYFGKSRATAERYIDGTNGGNTSNVRQLGKRSIKOTETLAROPLI 829  
Db 820 ELLGSRALNG-----NSRNTKATARTRPWYI 847



APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-Ping  
APPLICANT: Murdin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: Suite 701, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/474,671  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/337,483  
FILING DATE: 08-NOV-1994  
CLASSIFICATION: 530  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-465 MIS:vg  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 915 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-474-671-96

Query Match 5.7%; Score 278; DB 3; Length 915;  
Best Local Similarity 21.4%; Pred. No. 4.4e-15;  
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

QY 5 FRLKPICFYLMGVMLYHHSVAEDAGRAGSEAOIQVLEDVHVAK--RVPKDKKVFETDARA 62  
DB 7 FRNLICLSMTAL---PAVAENV-QAG-QAQEKQLDTIQVAKKQKTRDNEVTGLGKL 61  
QY 63 VSTRQDIFKSE-NLDNIVRSPG-AFTQDDK--SSGIVSINIRGDSGFGRVNTMVDGIT 118  
DB 62 VKTADTLKSEQVLDRLTRYDPCGIAVEQGRGASSG--YSIRG-MDKNRVSLTVGLA 117  
QY 119 Q-TPTYSTDAG--RAGSSQFQASVDSNFTAGLDVYKGFSSGAGINSAGSANLRTLG 175  
DB 118 QIOSYTAQAALGGTRTAGSGAINEIEYENVKAVEISKSGNSVBQSGGALAGSVAFQTKT 177  
QY 176 VDDVVGQNNYTGILLK-GLTGTNS--TKGNAMAA-ICARKWL--ESGASGVLYGHSRS 229  
DB 178 ADDVIGEGROWIQSTAYSGKNRGLTQSIALAGRIGGAELLIRTGHRHAGEIRAH---- 233  
QY 230 VAQNYRVGGGQHGIFGAELERRKORYFVQEGGLKFNNSGKWERDFORPYWKTWKYQ 289  
DB 234 -----EAGRGVQSFNRLAPVDDGSGYAYFIVEECK-----NGHEKCKANP----- 276

QY 290 KYNDPQELQKYEIGHDK-----SWRENLAPOYDIPTDPSSLKQOS-----AGNLF--KLEY 339  
DB 277 -----KKDVVGEDKQRTVSTRDYTGPNRFLA--DPLSYESRSWLFPRGCFRENRHY 326  
QY 340 DGVFNKYTAQ---FRDLNLTIGSRKII-----NRNQFNYGLSLNYSANLNL 383  
DB 327 IGGILERTQOTFTDRDMTVPAFLTKAVFDANQKQAGSLRNGNGKYAGNH-----KYGL-- 379  
QY 384 TAAYNSGRQYKPGSKFTGMLLKDFETYNNAKILDLNNTATFRLPRETELQTTLGFNYF 443  
DB 380 ---FTSGENNAAPVCAEY-GTGVFYD-----ETHKSRYGLEYV 413  
QY 444 HNEYGKNRFPPEELGLFFD--GPDQDNL-----YSYLRGFRFGDK 480  
DB 414 YTNADKDTWADYARLSYDRQDGLDNHFQOOTHCSADGSDKYCRPSADKPESY---YKSDR 470  
QY 481 GLLPOKSTIVQAGSOYFNTFYFDAALKDIYRLNYSNTVGY-RFGG-----EYTVGYG 534  
DB 471 VIYGESHKLQAAFKKSFTAKI-----RHNLNVN-LGYDRFGSNLRHODY--YYQ 518  
QY 535 SDEFFKRAFGENSPTYKKHCNQSCGIYEPVLKYGKKRA-----NNHSVYSISAD----- 583  
DB 519 SAN---RAYSLKTPP-----QNNGKKTSPNGREKNPYVWSIGRGNVTR 559  
QY 584 ---FGD-----YFMPEA-----SYSRTH----- 598  
DB 560 QICLFNGNTYTDCPTRSINGKSYAAVRDNVRLGRWADVGAGLRDYRSTHSDGVSSTG 619  
QY 599 -----RMPNIOEMVFSQIGDSGVHTALKPERRANTWQ 629  
DB 620 THRTLSWNAIVLKPADWLDLTYRTSTGFRLPSPAEMYGWSGDKIKAVKIDPEKSNKE 679  
QY 630 FG-----FNTYKGLLKQDDTLGLKLGVYRSRDN-----YIH--- 662  
DB 680 AGIVFKGDFGNLEASFNNAYRDLIVR-----GYEAOIKDGKEQVKNPAYLNAQS 730  
QY 663 -----NVYCK--W---WDLNGNIP-SWVSSTGLAYT-IQHRNEKDKVHKHGFLELNYD 709  
DB 731 ARITGINILKIDWNGVMD---KLPEGWYST--FAYNRVRVDIKKRADRTDIOSHLD-FD 784  
QY 710 YGREFTNLSYAYKSTQPTNFSDASESPNNAKEDQLKQGYGLSRVSALPRDYGRLEVGT 769  
DB 785 -----AIQPSRYV-----VCSGYDQPEGKGVNGMLT-----YSKAREIT 819  
QY 770 RWLGNKLTGGMRYFKGKSIKATAERYIDGTNGNTSNVRQLGKRSIKQETILARPLI 829  
DB 820 ELLGSRALNG-----NSRNTKATARRTRPWYI 847  
QY 830 FDFVAAVEPKKNLIFRAEVKNLFDRIYI 857  
DB 848 VDVSGYITVKKHFTLRAGVYNLLNHRVY 875

RESULT 8  
US-08-483-577A-96  
; Sequence 96, Application US/08483577A  
; Patent No. 6015688  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena  
; APPLICANT: Harkness, Robin  
; APPLICANT: Schryvers, Anthony  
; APPLICANT: Chong, Pele  
; APPLICANT: Gray-Owen, Scott  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Murdin, Andrew  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Transferrin Receptor Genes  
; NUMBER OF SEQUENCES: 160  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto

STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICANT: US/08/483,577A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/337,483  
FILING DATE: 08-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-511  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 915 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-483-577A-96

Query Match 5.7%; Score 278; DB 3; Length 915;

Best local similarity 21.4%; Pred. No. 4.4e-15;

Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

QY 5 FRKPCIFGVMGLYHSHSVAEDAGRAGSEAOIQVLEDVHVAK--RVPKDKKVFETDARA 62  
DB 7 FRMLNLCSLMTAL--PAAENV-QAQEQKQDITQVAKKQKTRDRNEVTGLGKL 61  
QY 63 VSTRQDIFKSSSE-NLDNIVRSIPG-AFTQDQK--SSGIVSLINIRGDSGFRVNTFMDGIT 118  
DB 62 VKTADTLKQVLDIRDITRYDPCIAVVEQGRGASSG--YSIRG-MDKNVRSLTVDGLA 117  
QY 119 Q-TPYSTSTAG--RAGSSQFASVDSNFIAGLDVVYKVSFSGSAGINSLAGSANLRTLG 175  
DB 118 QIQSYTAQAALGGTRTAGSSGAINIEIYENVKAVEISKNSVBOGSGALAGSVAFOTKT 177  
QY 176 VDDVVQGNNTYGLLLK-GLTGTSN--TKGNAMAA-IGARKWL--ESGASGVLYGHSSRS 229  
DB 178 ADDVIGSGROWIQSKTAYSGKNGLTQSLAGRIGAGBAILIRTHAGETRAH----- 233  
QY 230 VAQNYRVGGGQHGIFNGFABYLERRKQRYFVQEGGLFNSNSGKWERDFORPYWKTWKYQ 289  
DB 234 -----EAAGRGVQSFNRLAPVDDGSKAYFIVEECK-----NGHECKANP----- 276  
QY 290 KYNDPQELQYIEGHDK-----SWRENAPQYDITPIDPSLKKQOS-----AGNLF--KLEY 339  
DB 277 -----KKDVGVEDKQTVSTRTDYTGPNRFLA--DPLSYESRSWLFPRGFRFENKRHY 326  
QY 340 DGVENKYTAQ---PRDLNTRIGSKHII-----NRNVOFNGLSLNSYANLNL 383  
DB 327 IGGILETQOTFTDRDINVPVAFITKAVFDANQKQAGSLRNGKYAGNH-----KYGL-- 379  
QY 384 TAAYNSGRQYKPGSKFTGMLLKDFEETYNNAKILDLNNTATFRLPRETELQTTLGFNYF 443  
DB 380 ---FTSGENNAPVGAEE-GTGVEFD-----ETHTKSRYGLEYV 413

QY 444 HNEYKNRFPPEELGLFFD--GPDDNGL-----YSYLGFRFKGDK 480  
DB 414 YTNADKDTWADYARLSYDRQIGLDNHNFOQTHCSADSKYCRPSADKPSY---YKSDR 470  
QY 481 GLLPQKSTIVOPAGSQYFNTFYFDAALKKDIIYRLNYSNTVGY-RFGG-----EYTGYYG 534  
DB 471 VIYGESHKLLQAFAFKSFDATAI-----RHLNLSVN-LGYDRFGSLNRHODY--YYQ 518  
QY 535 SDDEFKRAFGENSPYKHKCNQSGIYEPVLKVKYKKRA-----NNHVSYSISAD----- 583  
DB 519 SAN---RAYSCLKTTP-----QNNGKKTSNPREKKNPYWYSIGRGNVVTR 559  
QY 584 ---FGD-----YFMPFA-----SYSRTH----- 598  
DB 560 QICLFCNNYTDCTPRSINKSYAAVRDNRVLGRWADVGAGLYDYRSTHSDDGSYSTG 619  
QY 599 -----RMPNQEMYFSGOISGVHTALKPRANTWQ 629  
DB 620 THRTLSWNAIGVLKPADWLDLTYRTSTGFRLPSPAEMYGWRSGDKIKAVKIDPEKSFKE 679  
QY 630 FG-----PNTYKKGILLKODDTLGLKLVCYRSIDN-----YIH--- 662  
DB 680 AGIVFKGDFGNLEASWFFNNAYRDLIVR-----GYEAQIKDGKEQVKGNPAYLNAQS 730  
QY 663 -----NVYKG--W---WDLNGNIP-SWVSSTGLAYT-IQHRNFKDKVHKHGFLELNYD 709  
DB 731 ARITGINILGKIDWNGVWD--KLPEGWYST--PAYNRVRVDIKKRADRTDIQSHL-FD 784  
QY 710 YGRFTNLAYOKSTOPTNFSDAESPNNASKEDQLKGGLSRVSALPRDVGRLVGT 769  
DB 785 -----AIOPSRYV-----VGSYDQPEGKVGNGMLT-----YSKAKEIT 819  
QY 770 RWLGNKLTGGAMRYFGKISRATAEERYIDGTNGNTSNVRQLGKRSIKOTETLAROLPI 829  
DB 820 ELLGSRALING-----NSRNTKATARTTRPYWI 847  
QY 830 FDYAAVEPKKNLIFRAEVKNLFDRIYI 857  
DB 848 VDVGSGYTVTKKHFTLRAGVYNLLNHRVY 875  
RESULT 9  
US-08-897-438-96  
; Sequence 96, Application US/08897438  
; Patent No. 6262016  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena  
; APPLICANT: Harkness, Robin  
; APPLICANT: Schryvers, Anthony  
; APPLICANT: Chong, Pele  
; APPLICANT: Gray-Owen, Scott  
; APPLICANT: Yang, Yan-ping  
; APPLICANT: Murdin, Andrew  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Transferrin Receptor Genes  
; NUMBER OF SEQUENCES: 160  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/897,438  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:



SEQUENCE CHARACTERISTICS:  
LENGTH: 915 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-637-654-96

Query Match 5.7%; Score 278; DB 4; Length 915;  
Best Local Similarity 21.4%; Pred. No. 4.4e-15;  
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

QY 5 FRLKPICFYLMGVMLYHSHYEDRAGRAGSEAOIQVLEDVHVHAK--RVPKDKKVFETDARA 62  
Db 7 FRLNLCLSLMTAL---PAYAENY-QAG-QAQRKQLODTIQVKKARKKQTRRDNEVTGLCKL 61  
QY 63 VSTRQDIFKSSSE-NLDNIVRSIPG-AFTQODK---SSGIVSLNIRGDSGFGRVNTWVGIT 118  
Db 62 VKTADTSLKEQVLDIRLDTRDPIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA 117  
QY 119 Q-TFYSTSTDAQ--RAGSGQFASVDSNFTAGLDVVKVFGSGAGINSLAGSANLRTLGL 175  
Db 118 QIQSYTAQAALGGTRTAGSGAINIEIYENVKAVEISKGSNSVEQSGSALAGSVAFOQTKT 177  
QY 176 VDDVVQGNNTYGLLLK-GLTGTNS--TKGNAMAA-IGARKWL--ESGASGVLYGHRSRS 229  
Db 178 ADDVIGGRQMGIOSTKATYSKNGRLTQSTIALAGRIGAEALLIRTGRHAGEIRAH---- 233  
QY 230 VAQNYRVGGGQHIGNFGAEYLERRKORYFYQEGGLKFNSNGKWERDFORPYWKTKWYQ 289  
Db 234 -----EAAGRGVQSFNRLAPVDDGSKYAFIVEECK-----NGHECKKANP----- 276  
QY 290 KYNDPQELQKYEIHKD-----SWRENLAQVDITPIDPSLKQOS-----AGNLF--KLEY 339  
Db 277 -----KKDVGEDKQRTVSTRDYTGPNRFLA--DPLSYESRSWLFRCGFRFENRHY 326  
QY 340 DGVFNKYTAQ---PRDLNLTIGSKII---NRYQFNYSGLNSYANLNL 383  
Db 327 IGGILERTQOTFDRTMTVPFLTKAVFDANOKQAGSLRGNGKYAGNH-----KYGL-- 379  
QY 384 TAAVNSGRQKPKSGKFTGMGLLKDFETNNYNAKILDLNNTATFRLPRETELQTTLGFNYF 443  
Db 380 ---FTSGENAPVGAEV-GTGVFYD-----ETHKSRYGLEYV 413  
QY 444 HNEYGKRNFPPELGLFED--GPDQDNGL-----YSYLRGRFKGDK 480  
Db 414 YTNADKDTWADYARLSYDROGIGLDNHFQOHCSDAGSDKYCRPSADKPFYSY---YKSDR 470  
QY 481 GLLPKSTIVOPAGSQVFNFTFYDAAALKDIIYRLNYSNTVGY-RFGG-----EYTCYIG 534  
Db 471 VIYGESHKLLQAPKKSFDTAI-----RHLNLSVN-LGVDRFGSNLRHODY--YYQ 518  
QY 535 SDDEFFKRAFGENSPYKKHCKNQSGGIYEPVLKKYKKRA-----NNHSVSIASAD----- 583  
Db 519 SAN---RAYSLKTPP-----QNNKKTSNGREKNPYWYISGRGNVTR 559  
QY 584 ---FGD-----TFMPFA-----SYSRTH----- 598  
Db 560 QICLFGNNTYDCTPRSINGKSYAAVRNRLGRWADVAGLRYDYRSTHSDDGSVSTG 619  
QY 599 -----RMPNIQEMVFSQIGDSGVHTALKPERANTWQ 629  
Db 620 THRTLSWNAIVLKPADWLDLTYRTSGFRLPSPFAEMYGWRSGDKIKAVKIDPEKSENKE 679  
QY 630 FG-----FNTYKKGILLKODDTLGLKLVGRSRIDN-----YIH--- 662  
Db 680 AGIVFKGDFCNLEASWFNNAYRLIVR-----GYEAIQDKGEQVKNPAYLNAQS 730  
QY 663 -----NVYQK--W---WDLGNIP-SWVSTGLAYT-IQHRNFKDKVHKHGFELNLYD 709  
Db 731 ARITGINILKIDWNGVWD---KLPEGWYST--FAYNRVRVRDIKKRADRTDIQSHL-FD 784  
QY 710 YGRFTNLVAYQKSTQPTNFSASESPNNAKEDQLKQGYGLSRVSALPRDYGRLEVGT 769

Db 785 -----AIQPSRYV-----VSGYDQPGKGVNGMLT-----YSKAKEIT 819  
QY 770 RWLGNKLTLCGAMRYFGKSIIRATAEERVIDGTNGNTSNVROLGRSIKQITETLARQPLI 829  
Db 820 ELLGSRALLNG-----NSRNTKATARRTRPWYI 847  
QY 830 FDFYAAAYEPKKNLIFRAEVKNLFDRIYI 857  
Db 848 VDVSGYTVVKKHFTLRAGVYNLLNHRVY 875

RESULT 11  
US-08-487-890A-94  
Sequence 94, Application US/08487890A  
Patent No. 5708149  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-ping  
APPLICANT: Murdin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,890A  
FILING DATE: 07-JUN-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-466 MTS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 908 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-487-890A-94

Query Match 5.5%; Score 269.5; DB 1; Length 908;  
Best Local Similarity 21.0%; Pred. No. 2.3e-14;  
Matches 216; Conservative 116; Mismatches 354; Indels 343; Gaps 49;  
QY 5 FRUKPICFYLMGVMLYHSHYEDRAGRAGSEAOIQVLEDVHVHAK--RVPKDKKVFETDARA 62  
Db 7 FRLNLCLSLMTAL---PYAENV--QAEQAQEKQDITQVKKAKKQKTRRDNEVTGLCKL 61



```
QY 235 RVGGGQHIGNFGAEYLERRKQ-----RYFVQE-----GGLKFNSN-----SGKWERDF 278
Db 234 KDAGKG--VQSFNRLVLDDEKKEGGSQRYRIFVEECHNGYAAACKNKLKEDASVKDER-- 289
QY 279 ORPYWKTKWQKYNDOE--LQKYIEGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKL 337
Db 290 -----KTVSTQDYTGSNRLLANPLEYGSQSWL--FRPGW-----HLDNR 326
QY 338 EYDGVENKYTAQFRDLNLTGSRKIIINRYOFNYSGLNSYANLNTAAYNSGRQKYPK- 396
Db 327 HYVGAVLERTQQTFTDRTMTVPAYFTSEDYVPGSLKGLGY-----SGDNKAERL 376
QY 397 -----GSKFTGWLKDFETYNNAKILDLNLTATFRLPRETELQTTGLGFNYFHEYKGNR 451
Db 377 FVQEGSTLOGIG-----YGTGVFYD-----ERHTKNRYGVVYVHNADKDT 418
QY 452 FPEELGLFFD--GPDQDNGL-----YSYLGFRFGDKGLLPKST 488
Db 419 WADYARLSYDRQIGIDLNRLOQTHCHSHGSDKNCRPDGNKPYSP---YKSDRMIYEESRN 475
QY 489 IVOPAGSQYFNT-----FYFDAAL----- 507
Db 476 LFOAVEKKAFTAKIRHNSLNGYDRFKSOLSHSDYLLQNAVOAYDLITPKPPFPNGS 535
QY 508 KKDYLRYNSTNVG-----YRFGGE-YT-----GYGSDDEPKRAFGENSPYKK 552
Db 536 KDNPRYVSIGKTYNTSPICRFGNTYTDCTPRNIGNGYV-----AAVDQNVRLGR 587
QY 553 HCNOSCGI-YEPLVKKYKGRANHSYS-----ISADFGDYFMPFASYSRTH-----R 599
Db 588 WADVAGAIRD-----YRSTHSEDKSVSTGTHRNLSNAGVVLKPFWTWMDLTYRASTGFR 642
QY 600 MPNTOEMFYFSGIDSGVHTLKPERANTWQFG-----PNTYKKGILLKQDDT 645
Db 643 LPSPAEMYGWRAGESLTKLDLKPESFNREAGIVFKGDFGNLEASYFNNAIRDLI----- 697
QY 646 LGLKLGYRSRIDN-----YIH-----NYGK--WMDLNGNIPSWSVSTGL 684
Db 698 ----AFGYETRTQNGTSASGDPYRNAQNAIAGINILGKIDWHGVWGLPDGLYST-L 752
QY 685 AY-----TIQHRNFQKVKHKGFELELYDY--GFFFTNLSYAYQKSTQPT 728
Db 753 AYNRIKVKDADIRDTFVTSYFADVQPSRYVLGLGYDHPDGIGWINTMTFTYSK----- 807
QY 729 NFSDAESPNNASKEDQKQGYGLSRVSALPRDYGRLEVTRWLGKNTLTLGGAMRYEFGKS 788
Db 808 -----AKSVDEL-----LGSQALLNG----- 823
QY 789 IRATAERYIDGTNGGNTSNVRLGKRSIKQTTETLARQPLIFDYAAAYEPKKNLIPRAEV 848
Db 824 -----NANAKAASRR-----TRPWYVTDVSGYINIKKHLTLRAGV 859
QY 849 KNLFDRIYI 857
Db 860 YNLLNRYIV 868

RESULT 13
US-08-337-483-94
; Sequence 94, Application US/08337483
; Patent No. 5922562
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
```

```
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-410 MIS:Jb
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-337-483-94
```

Query Match 5.5%; Score 269.5; DB 2; Length 908;  
Best Local Similarity 21.0%; Pred. No. 2.3e-14;

Matches 216; Conservative 116; Mismatches 354; Indels 343; Gaps 49;

```
QY 5 FRUKPCIFUMGVMVYHSHSAEDAGRAGSEAQVLEDVHVAK--RVPDKKVFVTPARA 62
Db 7 FRNLILCLSLMTAL---PYAENV--QAEQAQEKQLDTIQVAKKQKTRRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE--NLDNIVRSIPG-AFTODK--SSGIVSLNIRGDSGFRVNTMVDGIT 118
Db 62 VKSDDTLRSKQVLNIRDLTRYDPGIAVVEQGRASSG---YSIRG-MDKNRVSLTVDGVS 117
QY 119 Q-TFYSTSTAG--RAGGSQFGASVDNFIAGLDVVYKGSFGSAGINSLAGSANLRTLG 175
Db 118 QIOSYTAQAALGCTRTAGSSGAINIEYENVKAVEISKGSNSSEYNGALAGSVAEQTKT 177
QY 176 VDDVOGNNTYGLLLK--GLTGTNSTKGNMAAATGARKWLESAGSVGLYGHSSRSVAQNY 234
Db 178 AADIIEGKGQWQISQTAYSGKDHALTQSLALAGR---SGGAELLIYTKRGRETHAH 233
QY 235 RVGGGQHIGNFGAEYLERRKQ-----RYFVQE-----GGLKFNSN-----SGKWERDF 278
Db 234 KDAGKG--VQSFNRLVLDDEKKEGGSQRYRIFVEECHNGYAAACKNKLKEDASVKDER-- 289
QY 279 ORPYWKTKWQKYNDOE--LQKYIEGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKL 337
Db 290 -----KTVSTQDYTGSNRLLANPLEYGSQSWL--FRPGW-----HLDNR 326
QY 338 EYDGVENKYTAQFRDLNLTGSRKIIINRYOFNYSGLNSYANLNTAAYNSGRQKYPK- 396
Db 327 HYVGAVLERTQQTFTDRTMTVPAYFTSEDYVPGSLKGLGY-----SGDNKAERL 376
QY 397 -----GSKFTGWLKDFETYNNAKILDLNLTATFRLPRETELQTTGLGFNYFHEYKGNR 451
Db 377 FVQEGSTLOGIG-----YGTGVFYD-----ERHTKNRYGVVYVHNADKDT 418
QY 452 FPEELGLFFD--GPDQDNGL-----YSYLGFRFGDKGLLPKST 488
Db 419 WADYARLSYDRQIGIDLNRLOQTHCHSHGSDKNCRPDGNKPYSP---YKSDRMIYEESRN 475
QY 489 IVOPAGSQYFNT-----FYFDAAL----- 507
```





Db 643 LPSFAEMYGWRAGSLKTLDLKPEKSFNREAGIVFKDGFNLEASYNFNNAYRDLI----- 697  
QY 646 LGLKLVGRSRIDN-----YIH-----NVYK--WMDLNGNIPSWVSSTGL 684  
Db 698 ----AFGYETQNGQTSASCDPCYRNAQNARIAGINILCKIDHWGVWGLPDGLYST-L 752  
QY 685 AY-----TIQHRNFKDKVHKHGFLELNIDY--GRFFTNLSYAYOKSTQPT 728  
Db 753 AYNRIKVKDADIRADRTFVTSYLFDAVQPSRYVLGLGYDHPDGIWINTWFTYSK----- 807  
QY 729 NFSDASESPNNASKEDQLKQYGLSRVSALPRDYGRLEVGTRWLGKNTLGGAMRYEFGKS 788  
Db 808 ----AKSVDEL-----LGSQALLNG----- 823  
QY 789 IRATAERYIDGTNGGNTSVNROLGKRSIKQTEFLARQPLIFDFFYAAEYKKNLIFRAEV 848  
Db 824 ----NANAKKAASRR-----TRPWTVDSGYNNIKKHLTLRAGV 859  
QY 849 KNLFDRIYI 857  
Db 860 YNLLNRYV 868

RESULT 15

US-08-474-671-94  
; Sequence 94, Application US/08474671  
; Patent No. 6008326  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena  
; APPLICANT: Harkness, Robin  
; APPLICANT: Schryvers, Anthony  
; APPLICANT: Chong, Pele  
; APPLICANT: Gray-Owen, Scott  
; APPLICANT: Yang, Yan-ping  
; APPLICANT: Murdin, Andrew  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Transferrin Receptor Genes  
; NUMBER OF SEQUENCES: 147  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/474,671  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/337,483  
; FILING DATE: 08-NOV-1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/175,116  
; FILING DATE: 29-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/148,968  
; FILING DATE: 08-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-465 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 908 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-474-671-94  
  
Query Match 5.5%; Score 269.5; DB 3; Length 908;  
Best Local Similarity 21.0%; Pred. No. 2.3e-14;  
Matches 216; Conservative 116; Mismatches 354; Indels 343; Gaps 49;  
  
QY 5 FRLKPTCYFLMGVMLYHHSYAEDAGRAGSQAQIQVLEVDHVHKAK--RVPKDKKVFTDARA 62  
Db 7 FRNLILCLSLMTAL---PYVAENV--QAEQAQEQKLDITQVKAQKQKTRRDNEVTGLGKL 61  
QY 63 VSTRQDIFKSE--NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFCRVNTWVDGIT 118  
Db 62 VKSSDLSKEQVLLNIRDLTRYDFGIAVVEQGRGASS---YSIRG-MDKNRVSLTYDGV 117  
QY 119 Q-TFYSTSTDAG--RAGGSQFGASVDSNFIAGLDVVVKSGFSGSAGINSLAGSANLRTL 175  
Db 118 QIQSYTAQAALGGTRTAGSSGAINIEIYENVKAVEISKGSNSSEYNGALAGSVAFTKT 177  
QY 176 VDDVQGNNTYGLLLK--GLTGTNSTKGNMAAIGARKWLESASVGVLYGHSSRSVAQNY 234  
Db 178 AADIIGEGKQWGIQSKTAYSGKDHALTQSLALAGR---SGGAELLIYTKRGRRETHAH 233  
QY 235 RVGGGGQHIGNFGAEYLERRKQ-----RYFVQE--GGLKFNSEN-----SGKWERDF 278  
Db 234 KDAGKG--VQSFNRLVDEDKKEGGSQYRYFIVEEENHNGYAAACKNKLKEDASVKDER-- 289  
QY 279 QRPYWKTKYQKYNDOE--LQKYIEGHDKSWRENLAPOYDITPIDPSSLKQQSAGNLFKL 337  
Db 290 ----KTVSTQDYTGSNRLLANPLEYGSQSWL--FRGW-----HLDNR 326  
QY 338 EYDGVFNKYTAQPRDLNLTIGSKRIINRNQYFNYSLSNLANLTAAYNSGRQKYPK-- 396  
Db 327 HYVGAVLERTQQTFTDRMTVPAYFTSEYVPGSLKGLGKY-----SGDNKAERL 376  
QY 397 ----GSKFTGWLKLDKFETYNNAKILDLNLTATFRLPRETELQTLTGFFNYFNEYKCN 451  
Db 377 FVQEGSTLOGIG-----YGTGVFYD-----ERHTKNRYGVYVYHNADKDT 418  
QY 452 FPEELGLFFD--GPDQDNGL-----YSYLGRFKGDKGLLPKST 488  
Db 419 WADYARLSYDROGIDLDLNLQOQTHCSHDGSKNCRPDGNKPYSF---YKSDRMIEBSRN 475  
QY 489 IVOPAGSQYFNT-----FYDAAAL----- 507  
Db 476 LFOAVFKKAFDTAKIRHNLNLSINLGYSQSLSHSDYLLQNAVQAYDLITPKPPFPNGS 535  
QY 508 KKDIYRLNYSNTVG---YRFGE-YT-----GYGSDDEKPKAFGENSPYKK 552  
Db 536 KNPYRVSIGKTTVNTSPICRFGNNTYDCTPRNIGNGYI-----AAVDNVLGR 587  
QY 553 HCNQSCGI-YEPVLKRYKKRANNHSVS-----ISADFGDYMFPFASYSRTH-----R 599  
Db 588 WADVGAGIRYD---YRSTHSEDKSVSTGTHRNLSNAGVWLKPTWMDLTYRSTGFR 642  
QY 600 MPNIQMYFSQIGDSGVHTALKPERANTWQF-----FNTYKGLLQKDDT 645  
Db 643 LPSFAEMYGWRAGESLKTLDLKPESKSFNREAGIVFKDGFNLEASYNFNNAYRDLI----- 697  
QY 646 LGLKLVGRSRIDN-----YIH-----NVYK--WMDLNGNIPSWVSSTGL 684  
Db 698 ----AFGYETRTONGQTSASGDCPYRNAQNARIAGINILCKIDHWGVWGLPDGLYST-L 752  
QY 685 AY-----TIQHRNFKDKVHKHGFLELNIDY--GRFFTNLSYAYOKSTQPT 728  
Db 753 AYNRIKVKDADIRADRTFVTSYLFDAVQPSRYVLGLGYDHPDGIWINTWFTYSK----- 807  
QY 729 NFSDASESPNNASKEDQLKQYGLSRVSALPRDYGRLEVGTRWLGKNTLGGAMRYEFGKS 788

Db 808 -----AKSVDEL-----LGSOALLNG----- 823  
QY 789 IRATAERYIDGTNGNTSNVROLGKRSIKOTETLARQPLIFDFYAAAYEPKKNLIFRAEV 848  
Db 824 -----NANAKKASRR-----TRPWVVTDSGYINIKKHLTLRAGV 859  
QY 849 KNLFDREYI 857  
Db 860 YNLLNYRYV 868

Search completed: July 24, 2002, 08:53:48  
Job time: 520 sec





```
Qy 61 RAVSTRQDIFKSSSELDNIVRSIPGAFTQODKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
Db 61 RAVSTRQDIFKSSSELDNIVRSIPGAFTQODKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
Qy 121 FYSTSTADAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
Db 121 FYSTSTADAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
Qy 181 QGNNTYGLLLKGLTGTNSTKGNMAAICARKWLESASGVLYIGHSRSSVAQNTYRVGGG 240
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAICARKWLESASGVLYIGHSRSSVAQNTYRVGGG 240
Qy 241 OHIGNFGAEYLERKQRYFVQEGGLKFNSNGSKWERDFQRPYKTKWYQKYNQDPELOKY 300
Db 241 OHIGNFGAEYLERKQRYFVQEGGLKFNSNGSKWERDFQRPYKTKWYQKYNQDPELOKY 300
Qy 301 IEHDKSWRENLAPOYDITPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQFRDLNKTIGSR 360
Db 301 IEHDKSWRENLAPOYDITPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQFRDLNKTIGSR 360
Qy 361 KIINRNTQFNYSLSNYANLNTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILDL 420
Db 361 KIINRNTQFNYSLSNYANLNTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILDL 420
Qy 421 NNTATFRLPRETELOTTLGFNFYFNEHYGKNRPFPEELGLFDPGDDNGLYSYLGFRFGDK 480
Db 421 NNTATFRLPRETELOTTLGFNFYFNEHYGKNRPFPEELGLFDPGDDNGLYSYLGFRFGDK 480
Qy 481 GLLPQKSTIVOPAGSQVFNTPYFQAGLKDIIYRLNYSNTVGYRFGGEYTYGYSDDDEFK 540
Db 481 GLLPQKSTIVOPAGSQVFNTPYFQAGLKDIIYRLNYSNTVGYRFGGEYTYGYSDDDEFK 540
Qy 541 RAFGENSPTYKKHCNQSCGIYEPVLYKKGKRRANNHSSVSIADFGDYFMPFASYSRTHRM 600
Db 541 RAFGENSPTYKKHCNQSCGIYEPVLYKKGKRRANNHSSVSIADFGDYFMPFASYSRTHRM 600
Qy 601 PNIQEMYSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660
Db 601 PNIQEMYSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660
Qy 661 IHVYGGKWDNLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNLDYGRFFTNLSYA 720
Db 661 IHVYGGKWDNLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNLDYGRFFTNLSYA 720
Qy 721 YOKSTOPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGKLTGG 780
Db 721 YOKSTOPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGKLTGG 780
Qy 781 AMRYFGKSIIRATAERYIDGTNGNTSNVRLQKRSIKQETLARQPLIFDFYAAYPEPK 840
Db 781 AMRYFGKSIIRATAERYIDGTNGNTSNVRLQKRSIKQETLARQPLIFDFYAAYPEPK 840
Qy 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEEVTNCNADKTLCKNGYGGT 900
Db 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEEVTNCNADKTLCKNGYGGT 900
```

## RESULT 2

```
US-09-762-926-2
; Sequence 2, Application us/09762926
; GENERAL INFORMATION:
; APPLICANT: Thonnard, Joelle
; TITLE OF INVENTION: Novel Compounds
; FILE REFERENCE: BM45330
; CURRENT APPLICATION NUMBER: US/09/762, 926
; CURRENT FILING DATE: 2001-02-14
; PRIOR APPLICATION NUMBER: PCT/EP99/05989
; PRIOR FILING DATE: 1995-08-13
; PRIOR APPLICATION NUMBER: GB 9818004.5
```

```
; PRIOR FILING DATE: 1998-08-18
; NUMBER OF SEQ ID NOS: 9
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 922
; TYPE: PRT
; ORGANISM: Bacteria
US-09-762-926-2
```

```
Query Match 99.88; Score 4894; DB 21; Length 922;
Best Local Similarity 99.88; Pred. No. 0;
Matches 920; Conservative 1; Mismatches 1; Indels 0; Gaps 0;
```

```
Qy 1 MRSSFRLKPTICFYLMGVMLYHHSVAEDAGRAGSEAOQLVEDVHVAKRVPKDKKVFDTA 60
Db 1 MRSSFRLKPTICFYLMGVMLYHHSVAEDAGRAGSEAOQLVEDVHVAKRVPKDKKVFDTA 60
Qy 61 RAVSTRQDIFKSSSELDNIVRSIPGAFTQODKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
Db 61 RAVSTRQDIFKSSSELDNIVRSIPGAFTQODKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
Qy 121 FYSTSTADAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
Db 121 FYSTSTADAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180
Qy 181 QGNNTYGLLLKGLTGTNSTKGNMAAICARKWLESASGVLYIGHSRSSVAQNTYRVGGG 240
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAICARKWLESASGVLYIGHSRSSVAQNTYRVGGG 240
Qy 241 OHIGNFGAEYLERKQRYFVQEGGLKFNSNGSKWERDFQRPYKTKWYQKYNQDPELOKY 300
Db 241 OHIGNFGAEYLERKQRYFVQEGGLKFNSNGSKWERDFQRPYKTKWYQKYNQDPELOKY 300
Qy 301 IEHDKSWRENLAPOYDITPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQFRDLNKTIGSR 360
Db 301 IEHDKSWRENLAPOYDITPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQFRDLNKTIGSR 360
Qy 361 KIINRNTQFNYSLSNYANLNTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILDL 420
Db 361 KIINRNTQFNYSLSNYANLNTAAYNSGRQKYPKGSKFTGWGLLKDFETYNNAKILDL 420
Qy 421 NNTATFRLPRETELOTTLGFNFYFNEHYGKNRPFPEELGLFDPGDDNGLYSYLGFRFGDK 480
Db 421 NNTATFRLPRETELOTTLGFNFYFNEHYGKNRPFPEELGLFDPGDDNGLYSYLGFRFGDK 480
Qy 481 GLLPQKSTIVOPAGSQVFNTPYFQAGLKDIIYRLNYSNTVGYRFGGEYTYGYSDDDEFK 540
Db 481 GLLPQKSTIVOPAGSQVFNTPYFQAGLKDIIYRLNYSNTVGYRFGGEYTYGYSDDDEFK 540
Qy 541 RAFGENSPTYKKHCNQSCGIYEPVLYKKGKRRANNHSSVSIADFGDYFMPFASYSRTHRM 600
Db 541 RAFGENSPTYKKHCNQSCGIYEPVLYKKGKRRANNHSSVSIADFGDYFMPFASYSRTHRM 600
Qy 601 PNIQEMYSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660
Db 601 PNIQEMYSQIGDSGVHTALKPERANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660
Qy 661 IHVYGGKWDNLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNLDYGRFFTNLSYA 720
Db 661 IHVYGGKWDNLNGNIPSWSSSTGLAYTIOHRNFKDKVHKHGFELNLDYGRFFTNLSYA 720
Qy 721 YOKSTOPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGKLTGG 780
Db 721 YOKSTOPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGKLTGG 780
Qy 781 AMRYFGKSIIRATAERYIDGTNGNTSNVRLQKRSIKQETLARQPLIFDFYAAYPEPK 840
Db 781 AMRYFGKSIIRATAERYIDGTNGNTSNVRLQKRSIKQETLARQPLIFDFYAAYPEPK 840
Qy 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEEVTNCNADKTLCKNGYGGT 900
Db 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEEVTNCNADKTLCKNGYGGT 900
```

QY 901 SKSVLTNFARGRTFLITMSYKF 922  
Db 901 SKSVLTNFARGRTFLITMSYKF 922

RESULT 3  
US-09-762-926-6  
; Sequence 6, Application US/09762926  
; GENERAL INFORMATION:  
; APPLICANT: Thonnard, Joelle  
; TITLE OF INVENTION: Novel Compounds  
; FILE REFERENCE: BM45330  
; CURRENT APPLICATION NUMBER: US/09/762,926  
; PRIOR FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: PCT/EP99/05989  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: GB 9818004.5  
; PRIOR FILING DATE: 1998-08-18  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 6  
; LENGTH: 921  
; TYPE: PRT  
; ORGANISM: Bacteria  
US-09-762-926-6

Query Match 97.5%; Score 4779.5; DB 21; Length 921;  
Best Local Similarity 97.9%; Pred. No. 0;  
Matches 903; Conservative 7; Mismatches 11; Indels 1; Gaps 1;

QY 1 MRSSFLKPCFCYLMGYLHSHYAEADAGRAGSEAIQVLEDVHVAKRVPKDKKVFDA 60  
Db 1 MRSSFLKPCFCYLMGYLHSHYAEADAGRAGSEAIQVLEDVHVAKRVPKDKKVFDA 60

QY 61 RAVSTRODIEKSSNLDNIIVRSIPGAFTQODKSSGIVSLNIRGDSGGRVNTWVDGITOT 120  
Db 61 RAVSTRODIEKSSNLDNIIVRSIPGAFTQODKSSGIVSLNIRGDSGGRVNTWVDGITOT 120

QY 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180  
Db 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180

QY 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYCHSRSSVAQNYRVGGG 240  
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYCHSRSSVAQNYRVGGG 240

QY 241 OHICNFAEYLERRKORYFVQEGGLKFNNSGKWERDFORPYKTKWYKYNDOPELOKY 300  
Db 241 OHICNFAEYLERRKORYFVQEGGLKFNNSGKWERDFORPYKTKWYKYNDOPELOKY 300

QY 301 IEHDKSWRENLAPOYDITPIDPSSLKQSQAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 360  
Db 301 IEHDKSWRENLAPOYDITPIDPSSLKQSQAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 360

QY 360 KIINRNTQFYGLSLNPTNLNTAAVNSGRQKYPKSKFTGWLKDKFTYNNAKLTLDL 419  
Db 360 KIINRNTQFYGLSLNPTNLNTAAVNSGRQKYPKSKFTGWLKDKFTYNNAKLTLDL 419

QY 421 NNTATFRLPRETLOTTFLGFNHEYGKRFPEELGLFDDGDDOGLSYLGRPKGDK 480  
Db 421 NNTATFRLPRETLOTTFLGFNHEYGKRFPEELGLFDDGDDOGLSYLGRPKGDK 480

QY 481 GLLPQKSTIVQAGSOFNTFYFDDAALKDIIYRLNYSNTVYRFGGEYTYGYSDDDEFK 540  
Db 481 GLLPQKSTIVQAGSOFNTFYFDDAALKDIIYRLNYSNTVYRFGGEYTYGYSDDDEFK 540

QY 541 RAFGENSPYKHKCNQSCGIIYEPVLYKYGKRRANNHVSISADFGDYFMPFASYSRTHRM 600  
Db 541 RAFGENSPYKHKCNQSCGIIYEPVLYKYGKRRANNHVSISADFGDYFMPFASYSRTHRM 600

QY 601 PNIQEMYFSQIGDSGVHTALKPERANTWQFCFNTYKKGLLKQDDTLGLKLVGYSRIDNY 660

Db 600 PNIQEMYFSQIGDSGVHTALKPERANTWQFCFNTYKKGLLKQDDTLGLKLVGYSRIDNY 659  
QY 661 IHNVYKQWDLNGNIPSWVSSTGLAYTIQHRNFKDKVHKHGFLELNYDYGRFFTNLSYA 720  
Db 660 IHNVYKQWDLNGDIPSWVSSTGLAYTIQHRNFKDKVHKHGFLELNYDYGRFFTNLSYA 719

QY 721 YOKSTQPTNFSASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWGLNKLTLGG 780  
Db 720 YOKSTQPTNFSASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWGLNKLTLGG 779

QY 781 AMRYFGKSIATAEERYIDCTNGGNTSNVROLGKRSIKQETTLARQPLIFDFYAYEPKK 840  
Db 780 AMRYFGKSIATAEERYIDCTNGGNTSNVROLGKRSIKQETTLARQPLIFDFYAYEPKK 839

QY 841 NLIFRAEVKNLDFDRYIDPLDAGNDAATQRYSSDFDKDEEVTNCNADKTLCKNGYGGT 900  
Db 840 NLIFRAEVKNLDFDRYIDPLDAGNDAATQRYSSDFDKDEEVTNCNADKTLCKNGYGGT 899

QY 901 SKSVLTNFARGRTFLITMSYKF 922  
Db 900 SKSVLTNFARGRTFLITMSYKF 921

## RESULT 4

US-09-303-518D-884  
; Sequence 884, Application US/09303518D  
; GENERAL INFORMATION:  
; APPLICANT: Scarlato, Vincenzo  
; APPLICANT: Masignani, Vega  
; APPLICANT: Rappuoli, Rino  
; APPLICANT: Pizza, Mariagrazia  
; APPLICANT: Grandi, Guido  
; TITLE OF INVENTION: Neisserial Antigens  
; FILE REFERENCE: CHIR0160  
; CURRENT APPLICATION NUMBER: US/09/303,518D  
; CURRENT FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 1098  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 884  
; LENGTH: 922  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-09-303-518D-884

Query Match 96.9%; Score 4754; DB 17; Length 922;  
Best Local Similarity 96.3%; Pred. No. 0;  
Matches 888; Conservative 16; Mismatches 18; Indels 0; Gaps 0;

QY 1 MRSSFLKPCFCYLMGYLHSHYAEADAGRAGSEAIQVLEDVHVAKRVPKDKKVFDA 60  
Db 1 MRSSFLKPCFCYLMGYLHSHYAEADAGRAGSEAIQVLEDVHVAKRVPKDKKVFDA 60

QY 61 RAVSTRODIEKSSNLDNIIVRSIPGAFTQODKSSGIVSLNIRGDSGGRVNTWVDGITOT 120  
Db 61 RAVSTRODIEKSSNLDNIIVRSIPGAFTQODKSSGIVSLNIRGDSGGRVNTWVDGITOT 120

QY 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180  
Db 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVV 180

QY 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYCHSRSSVAQNYRVGGG 240  
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYCHSRSSVAQNYRVGGG 240

QY 241 OHICNFAEYLERRKORYFVQEGGLKFNNSGKWERDFORPYKTKWYKYNDOPELOKY 300  
Db 241 OHICNFAEYLERRKORYFVQEGGLKFNNSGKWERDFORPYKTKWYKYNDOPELOKY 300

QY 301 IEHDKSWRENLAPOYDITPIDPSSLKQSQAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 360  
Db 301 IEHDKSWRENLAPOYDITPIDPSSLKQSQAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 360

QY 361 KIIIRNYQFNYGLSLNSYANLNTAAYNNSGRQYPKGSKFTGWGLLKDFETYNNAKILDL 420  
 Db 361 KIIIRNYQFNYGLSLNPYTNLNTAAYNNSGRQYPKGAKFTGWGLLKDFETYNNAKILDL 420  
 QY 421 NNTATFRLPRETELQTTFLGNYFHNEVGKRRPEELGLFFDGPDDONGLYSYLGRFKGDK 480  
 Db 421 NNTATFRLPRETELQTTFLGNYFHNEVGKRRPEELGLFFDGPDDONGLYSYLGRFKGDK 480  
 QY 481 GLLPQKSTIVOPAGSOFNTFYFDAALKKDIYRLNSTNTVGYRFGEGYTYGYGSDDEFFK 540  
 Db 481 GLLPQKSTIVOPAGSOFNTFYFDAALKKDIYRLNSTNTVGYRFGEGYTYGYGSENEFFK 540  
 QY 541 RAFGENSPYKKHCNOSCGIYEPVLKKYKKRANNSHSVISADFGDYFMPFASYSRTHRM 600  
 Db 541 RAFGENSPAYKEHCDPSCGIYEPVLKKYKKRANNSHSVISADFGDYFMPFAGYSRTHRM 600  
 QY 601 PNIQEMYFSOIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLGVYRSRIDNY 660  
 Db 601 PNIQEMYFSOIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLGVYRSRIDNY 660  
 QY 661 IHVNYGKWDNLGNIPIFSWSSSTGLATYIOHRNFKDKVHKHGELELUNDYGRFFTNLSYA 720  
 Db 661 IHVNYGKWDNLGDIPIFSWGSTGLATYIHRNFKDKVHKHGELELUNDYGRFFTNLSYA 720  
 QY 721 YOKSTQPTNFSDESPPNNAKEDQLKQGVLSRVLSALPRDYGRLEVGRWLGNKLTILGG 780  
 Db 721 YOKSTQPTNFSDESPPNNAKEDQLKQGVLSRVLSALPRDYGRLEVGRWLGNKLTILGG 780  
 QY 781 AMRYFGKSIRATAERYIDTNGGNTSNVRQLGKRSIKOTETLARQPLIFDFYAAAYEPKK 840  
 Db 781 AMRYFGKSIRATAERYIDTNGGNTSNVRQLGKRSIKOTETLARQPLIFDFYAAAYEPKK 840  
 QY 841 NLIFRAEVKNLFRRYIDPLDAGNDAATQRYSSFDPKDKDEVICNADKTLGNGKYGGT 900  
 Db 841 NLIFRAEVKNLFRRYIDPLDAGNDAATQRYSSFDPKDKDEVICNADKTLGNGKYGGT 900  
 QY 901 SKSVLTNFAFGRTFLTMSYKF 922  
 Db 901 SKSVLTNFAFGRTFLTMSYKF 922

RESULT 5  
 US-09-303-518D-882  
 ; Sequence 882, Application US/09303518D  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Scarlato, Vincenzo  
 ; APPLICANT: Massignani, Vega  
 ; APPLICANT: Rappuoli, Rino  
 ; APPLICANT: Pizza, Mariagrazia  
 ; APPLICANT: Grandi, Guido  
 ; TITLE OF INVENTION: Neisserial Antigens  
 ; FILE REFERENCE: CHIR0160  
 ; CURRENT APPLICATION NUMBER: US/09/303,518D  
 ; CURRENT FILING DATE: 1999-04-30  
 ; NUMBER OF SEQ ID NOS: 1098  
 ; SOFTWARE: PatentIn version 3.1  
 ; SEQ ID NO 882  
 ; LENGTH: 922  
 ; TYPE: PR1  
 ; ORGANISM: Neisseria gonorrhoeae  
 US-09-303-518D-882

Db	61	RAVSTRQDVAFSGENLDNIYRSIPGAFTQDKSSGIVSLNIRKDSGFRVNTWDGIIQT	120
Qy	121	FYSTSTDAGRAGGSSQFCASVDSNFITAGLDVWKGFSFGSAGINSLAGSANLRTLGVDVV	180
Db	121	FYSTSTDAGRAGGSSQFCASVDSNFITAGLDVWKGFSFGSAGINSLAGSANLRTLGVDVV	180
Qy	181	QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLKESGASVGLYGHSSRRAVQAQNYRVGGG	240
Db	181	QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLKESGASVGLYGHSSRRAVQAQNYRVGGG	240
Qy	241	QHIGNFAEYLERKQRYFVQEGGLFNSNSGKWERDFORPFWKTKWYQKYNDOPELQKY	300
Db	241	QHIGNFEEYLERKQDYFVQEGGLFNFAGSGKWERDLQRYWTKTKWYKKYEDPELQKY	300
Qy	301	IEGHDKSWRENLAPOYDITPIDPSSLLKQOSAGNLFKLEYDGVFNKYTAQFRLNTKIGSR	360
Db	301	IEEHDKSWRENLAPOYDITPIDPSGLKQOSAGNLLNLEYDGVFNKYTAQFRLNTFRIGSR	360
Qy	361	KIINRNTQFNYGLSLNYSANLNTAAYNRSGRQYKPGSKFTGWGLLKDPETYNNAKILDL	420
Db	361	KIINRNTQFNYGLSLNYPYTNLNTAAYNRSGRQYKPGAKFTGWGLLKDPETYNNAKILDL	420
Qy	421	NNTATFRLPRETELOTTLGFENYFHNKRPPEELGLFEFDGPDQDNGLYSYLGRFKGDK	480
Db	421	NNTATFRLPRETELOTTLGFENYFHNKRPPEELGLFEFDGPDQDNGLYSYLGRFKGDK	480
Qy	481	GLLPKQSTIVQAGSQYFNTFYFDDAALKKDIYRLNTSTNTVYRFGEGYTYGYSDDDFK	540
Db	481	GLLPKQSTIVQAGSQYFNTFYFDDAALKKDIYRLNLTNAINYRFGEGYTYGYSSENEFK	540
Qy	541	RAFGENSPYKHKCNOSCGIYEPVLKYYGKKRANNSHSVISADFGDYMFPFAGYSRTHRM	600
Db	541	RAFGENSPAYKEHCDPSCGLYEPVLKYYGKKRANNSHSVISADFGDYMFPFAGYSRTHRM	600
Qy	601	PNIOEMFYSGIDSGVHTALKPERANTWQGFNTYKKGLLKODDTLGLKLGVYRSRIDNY	660
Db	601	PNIOEMFYSGIDSGVHTALKPERANTWQGFNTYKKGLLKODDTLGLKLGVYRSRIDNY	660
Qy	661	IHNYYGKWDNLNGNIPSWSSSTGLAVTIQHRNFKDKVHKHGFLELELNDYGRFFFNLSYA	720
Db	661	IHNYYGKWDNLNDIPIWSWGSTGLAVTIIRHNFKDKVHKHGFLELELNDYGRFFFNLSYA	720
Qy	721	YOKSTQPTNFSDAESPNNASKEDQLKQYGLSRVSALPRDYGRLEVGTWLGKNTLGG	780
Db	721	YOKSTQPTNFSDAESPNNASKEDQLKQYGLSRVSALPRDYGRLEVGTWLGKNTLGG	780
Qy	781	AMRYFGKSIATAEERYIDGTNGNTSNVRQLGKRSIKOTETLAKOPLIFDYYAAYEPKK	840
Db	781	AMRYFGKSIATAEERYIDGTNGNTSNVRQLGKRSIKOTETLAKOPLIFDYYAAYEPKK	840
Qy	841	NLIFRAEYKVLNFORRYIDPLDAGNDAATQORYYSSDFPKDKDEVTCTNADKTLCLNGKYGT	900
Db	841	NLIFRAEYKVLNFORRYIDPLDAGNDAATQORYYSSDFPKDKDEVTCTNADKTLCLNGKYGT	900
Qy	901	SKSVLTNFAGRGTFLTMSYKF 922	
Db	901	SKSVLTNFAGRGTFLTMSYKF 922	

RESULT 6

US-09-303-518D-878

; Sequence 878, Application US/09303518D

; GENERAL INFORMATION:

; APPLICANT: Scarlato, Vincenzo

; APPLICANT: Masignani, Vega

; APPLICANT: Rappuoli, Rino

; APPLICANT: Pizza, Mariagrazia

; APPLICANT: Grandi, Guido

; TITLE OF INVENTION: Neisserial Antigens

; FILE REFERENCE: CHIR0160

; CURRENT APPLICATION NUMBER: US/09/303,518D

; CURRENT FILING DATE: 1999-04-30



```

; NUMBER OF SEQ ID NOS: 1098
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 878
; LENGTH: 888
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (279)..(279)
; OTHER INFORMATION: Xaa= any amino acid
US-09-303-518D-878

Query Match          93.8%; Score 4599.5; DB 17; Length 888;
Best Local Similarity 97.8%; Pred. No. 0;
Matches 869; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

QY 34 EAQIOVLEHVHAKRVPKDKVFTDARAVSTRODIFKSSSENLDNIVRSIPGAFTQODKS 93
Db 1 EAQIOVLEHVHAKRVPKDKVFTDARAVSTRODIFKSSSENLDNIVRSIPGAFTQODKS 60

QY 94 SGIVSLNIRGDSGFRVNTMVDGTTQTFTYSTDAGRAGSSQFGASVDSNFIAGLDVVK 153
Db 61 SGIVSLNIRGDSGFRVNTMVDGTTQTFTYSTDAGRAGSSQFGASVDSNFIAGLDVVK 120

QY 154 GSPSGSAGINSLAGSANLRITLGVDDVVQGNNTYGLLLKGLTGNTSTKGNMAAIGARKWL 213
Db 121 GSPSGSAGINSLAGSANLRITLGVDDVVQGNNTYGLLLKGLTGNTSTKGNMAAIGARKWL 180

QY 214 ESGASVGLVGHRSRVAQNYRVGGGGOHIGNFGAEYLERRKORYFVQEGGLFNSNSGK 273
Db 181 ESGASVGLVGHRSRVAQNYRVGGGGOHIGNFGAEYLERRKORYFVQEGALKFNSDSGK 240

QY 274 WERDFQRYWTKWKYQKYNQDPOELQKYIEGDKSWRENLAPOYDITPIDPSSLKQOSAGN 333
Db 241 WERDLQWQWKKYKPYKNYN-QELQKYIEHDKSWRENLAPOYDITPIDPSSLKQOSAGN 299

QY 334 LFKLEYDGVFNKYTAQPRDLNTKIGSRKIIIRNRYQFNYSLSANLNTAAYNSGRQK 393
Db 300 LFKLEYDGVFNKYTAQPRDLNTKIGSRKIIIRNRYQFNYSLSANLNTAAYNSGRQK 359

QY 394 YPKGSKFTGMLLKDFEYNNNAKILDNNATFLPRETELQTLTGFNYFNEYGKNRFP 453
Db 360 YPKGSKFTGMLLKDFEYNNNAKILDNNATFLPRETELQTLTGFNYFNEYGKNRFP 419

QY 454 EELGLFPDGDQDNGLYSLGRKFGDKGLLPQKSTIVOPAGSQYFNTFYDAALKKDIYR 513
Db 420 EELGLFPDGDQDNGLYSLGRKFGDKGLLPQKSTIVOPAGSQYFNTFYDAALKKDIYR 479

QY 514 LNYSTNTVGYRFGGEXTGYGSDDEKFRAGENSPTKKHCNQCGLIYEPVLKKYGGKRA 573
Db 480 LNYSTNTVGYRFGGEXTGYGSDDEKFRAGENSPTKKHCNQCGLIYEPVLKKYGGKRA 539

QY 574 NNHSVSIADFGDYFMPFASYSTRHPNPIQEMYFSGISGDSGVHTALKPERANTWQGFN 633
Db 540 NNHSVSIADFGDYFMPFASYSTRHPNPIQEMYFSGISGDSGVHTALKPERANTWQGFN 599

QY 634 TYKKGLLKQDDTLGLKLVGRSRIDNTYIHNVYKWDNLGNIPSWSSSTGLAYTIQHRNF 693
Db 600 TYKKGLLKQDDTLGLKLVGRSRIDNTYIHNVYKWDNLGNIPSWSSSTGLAYTIQHRNF 659

QY 694 KDKVHKHGFLELNVDYGRFFTNLSAYOKSTOPTNFSDASESPNNASKEDQLKQGYLS 753
Db 660 KDKVHKHGFLELNVDYGRFFTNLSAYOKSTOPTNFSDASESPNNASKEDQLKQGYLS 719

QY 754 RVSAIPRDYGRLEVGTGRLGNKLTGLGAMRYFGKSRATAEERYIDGTNGNTSNVROLG 813
Db 720 RVSAIPRDYGRLEVGTGRLGNKLTGLGAMRYFGKSRATAEERYIDGTNGNTSNFROLG 779

QY 814 KRSIKQETLARQPLIFDFYAAEYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYS 873
Db 780 KRSIKQETLARQPLIFDFYAAEYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYS 839
```

```

QY 874 SFDPKDKDEVTNADKTLNCGKYGGTSKSVLTNFAARGRTFLITMSYKF 922
Db 840 SFDPKDKDEVTNADKTLNCGKYGGTSKSVLTNFAARGRTFLITMSYKF 888

RESULT 7
US-09-303-518D-880
; Sequence 880, Application US/09303518D
; GENERAL INFORMATION:
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisserial Antigens
; FILE REFERENCE: CHIRO160
; CURRENT APPLICATION NUMBER: US/09/303,518D
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 1098
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 880
; LENGTH: 871
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (22)..(22)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (31)..(31)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (38)..(38)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (40)..(40)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (45)..(45)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (52)..(52)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (68)..(68)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (95)..(95)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (124)..(124)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (133)..(133)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (277)..(277)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (355)..(355)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (370)..(370)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (414)..(414)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (477)..(477)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (483)..(483)
; OTHER INFORMATION: Xaa= any amino acid
```



```
Db 407 VLAHNVGQYKPKGSTFTGWLKDKDFETKNTANFLDLNTHFTNLPKQMDLTTTGLNI 466
QY 443 FHNEYGNRRPEELGDFDGDQDNGLYSL-GRFGDKGLLPQKSTIVPAGSQYENTF 501
Db 467 LHNEYSKNRFPDELGLFTYNDLLCGGQYDAGCGRFQCTSTLPPKSVIVQPSGKQREHSI 526
QY 502 YFDAALKDIYRLNYSNTVGYRFGGEYTYGGYSDDEKFRAGFENSTYKHKHCNQCIGY 561
Db 527 YLDTSLQDKYQLDYSYNASQYRFSGBHASYYSQKQFQDKFCBDSQIYKQHCSPSCDVY 586
QY 562 EPLVKYKGRKRANHSYSISADFGDYEMPFASYSRTSRMNPNIQEMYSQIGDSGVHTALK 621
Db 587 EPLVTTSGKKHAINHSVTLAKYDTGPMFPVFSFARTSRMNPNIQEMFFSQIGDGVNTALK 646
QY 622 PERANTWOFQNTYKGLLKQDDTLGLKLVYGRSRIDNYIHNVYKWMDLNGNIPSWSS 681
Db 647 PEQANTYQLGFNFKRLNLLTDNDTLGLKVVGYQSRINNYIHNVYKWKYDTK-NPPSWVTS 705
QY 682 TGL-AYTIQHRNFKDKYKHKHGFLELNYDYGREFTNLSYAYQKSTQPTNFSASESPNNA 740
Db 706 GALKGDTIQHRNQWQMPVHKOGLELEINYGARYETNLSYARQKTDQPTNYSASESPNNS 765
QY 741 SKEDOLQGGYGLSRVSALPRDYGRLEVYGRWLGKGLTGGAMRYFGKSIIRATAERYIDG 800
Db 766 SKEDOLTQGGYGLSKVSMPLPRDYGRFELGVRCFDKLTIGSAVRYGQSPRATIEPRIDG 825
QY 801 TNGNTSNVROLGKRSTKQTTETLARQPLIFDYAAYPEKKNLIFRAEVKNLFDRIYIDPL 860
Db 826 THGNTSHSDDKGAHVIKQIEMLRQPLVDFYVAYEPIKDLVNRLDVQNAFDKLYIDPL 885
QY 861 DAGNDAATORYSSFDPKDDEEVTCKNADKTLG--NGKYGTSKSVLTNFARGTFLITM 918
Db 886 DANDAATORYHSI-YNDADEGAPCAAGQ-LCKPDAKYGGTTTSRLTNFAKGRSLLFSM 943
QY 919 SYKF 922
Db 944 TYKW 947

RESULT 9
US-09-303-518D-876
; Sequence 876, Application US/09303518D
; GENERAL INFORMATION:
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Massignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisserial Antigens
; FILE REFERENCE: CHIR0160
; CURRENT APPLICATION NUMBER: US/09/303,518D
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 1098
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 876
; LENGTH: 393
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (21)..(21)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (104)..(104)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (163)..(163)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (165)..(165)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (171)..(178)
```

```
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (301)..(301)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (339)..(339)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (353)..(353)
; OTHER INFORMATION: Xaa= any amino acid
US-09-303-518D-876

Query Match 40.3%; Score 1976; DB 17; Length 393;
Best Local Similarity 94.1%; Pred. No. 1.7e-175;
Matches 369; Conservative 6; Mismatches 17; Indels 0; Gaps 0;

QY 531 GYYGSDDEKFRAGFENSTYKHKHCNQCIGYEPVLKYYGKKRANHSYSISADFGDYFMP 590
Db 2 GYYGSDDEKFRAGFENSTYKHKHCNQCIGYEPVLKYYGKKRANHSYSISADFGDYFMP 61
QY 591 FASYSRTRHPNPIQEMYSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKL 650
Db 62 FASYSRTRHPNPIQEMYSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKL 121
QY 651 VGYRSRIDNYIHNVYKWMDLNGNIPSWVSTGLAYTIQHRNFKDKYKHKHGFLELNYDY 710
Db 122 VGYRSRIDNYIHNVYKWMDLNGDIPSWVSTGLAYTIQHRNFKDKYKHKHGFLELNYDY 181
QY 711 GRFTNLSYAYQKSTQPTNFSASESPNNAKEDQKGYGLSRVSALPRDYGRLEVYGR 770
Db 182 GRFTNLSYAYQKSTQPTNFSASESPNNAKEDQKGYGLSRVSALPRDYGRLEVYGR 241
QY 771 WLGKGLTGGAMRYFGKSIIRATAERYIDGTNGNTSNVROLGKRSTKQTTETLARQPLIF 830
Db 242 WLGKGLTGGAMRYFGKSIIRATAERYIDGTNGNTSNVROLGKRSTKQTTETLARQPLIX 301
QY 831 DFYAAYPEKKNLIFRAEVKNLFDRIYIDPLDAGNDAATORYSSFDPKDDEEVTCKNADK 890
Db 302 DFNAAYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAAXERYSSFDPKDDEEVTCKNADK 361
QY 891 TLCNGKYGTSKSVLTNFARGTFLITMSYKF 922
Db 362 TLCNGKYGTSKSVLTNFARGTFLITMSYKF 393

RESULT 10
US-09-543-681A-8096
; Sequence 8096, Application US/09543681A
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8096
; LENGTH: 760
; TYPE: PRT
; ORGANISM: Proteus mirabilis
US-09-543-681A-8096

Query Match 15.4%; Score 753.5; DB 19; Length 760;
Best Local Similarity 25.6%; Pred. No. 1.9e-60;
Matches 240; Conservative 144; Mismatches 328; Indels 225; Gaps 32;

QY 11 CFYLMGVNLYHHS--YAEDAGRAGSEAOIQVLEDVHVKAKEVPKDKVFTDARAVSTFQD 68
Db 24 CILFSGI-----HSLVFAAETKTVQFSLKV-----SSAQKQTPQKALSKPGAYSAGE 74
```

Qy	69	IFKSSENLDNIVRIPGAFTOQDKSSITVSLNRIGDSGRVNMVDGTOTFYSTSDA	128
Dd	70	: : : :       :       :	
Dd	75	I - NNLSSVEQALRSTPGTYTQMDSOPGVGNIRGLSGFRVNMMDGTWTMYSTSPSQ	133
Qy	129	GRAGS--SOGFASVDSNFAGLDVVYKGSFGSAGINSLAGSANLFTLGVDVVOGNNTY	186
Dd	130		
Dd	134	YAHGGQYNQFNMSIDPNFIQDISRGQODGENSNALAGSANFTIGDIDLFSNHKW	193
Qy	187	GLLLGLTGTSNKGNMAAI - GARKWLESAGSVLYGHRSRRSAQNRYVGGOHIGN	245
Dd	188	:	
Dd	194	GIRSKAARGTNGLYNGWVAIAKQPINFHDGYIGAMFALSCHNISSYKNAG----	247
Qy	246	FBAEYLERRKQTFVBEGGLKFNSNGKKWERDP- QRPYW - TKWYOKYNDPOELQKYIBG	303
Dd	248		
Qy	304	HDKSWRENLAPOYDITPIDPSSLKQOSAGNFLKLEYDVGFNYTAQFDRLNTIKGSRKI	363
Dd	305	: : : :       :       :       :       :       :	
Qy	364	NNRYQNYGLSLSYANLNLTAAYSNGROKYPKSGKFTGMGLLKDPETYNNAKILDLNNT	423
Dd	365	: : :     :     :     :     :     :     :     :	
Qy	424	ATFLPRETELQTTLGFNYPHNEYGNRRPEELGLFFDGDQDNGLYSYLGRFKGDKGLL	483
Dd	425	:   :	
Dd	355	SRENY - HELDMALTGLTKLSIDYNR-----EVTAPSDPWOSQNNIEY-----	397
Qy	484	PKSTIVQPSAGSVYNTFYFDAALKXDIYRLNYSYTVGYREGGEYTYGYSDDFEKRAF	543
Dd	485	:	
Dd	398	----NVFAPOGKDINSLSQMFEDITYLALUNLYTDYHKG-----	437
Qy	544	GENSPYKKHCNSCGIYEPVLKKYKKKRANNHSV-----ISADFGDFMPFPASYST	597
Dd	545	:	
Dd	438	-----YKPACDPKAACFPE-----GAMNVNRHEKAWEPGALFSAQIIPEFEPFISYAHT	486
Qy	598	HRMNPQEMFYSIGDSGVHTALKPERANTWQGFNTYKGLLKQDDTLGLKLGVYRSRI	657
Dd	599		
Dd	487	TRAPNQOEIFFANEGGASMPFLURSEKADTFQIGFNSYRPLDIVAGDSFLRKALMYHTKV	546
Qy	658	DNYT-----HNWY--GWWDMLNGNIP-----SWVSSTGLATTIOHRNEKDKVKHGFELELN	707
Dd	659		
Dd	547	KNYISSDYSYNVCKGGRCKIDGTTDDDYSGMEYDCNLYL--YTNSLDPVTVMRGYELQVN	604
Qy	708	YDYGRTFNLSUYAKYSTOPTNESDASESPNNAKEDQLKQG-YGLSRVSVALPRDYGRUE	766
Dd	709	:	
Dd	605	YDADVFTTLYSNSNERTSQPTS-----LAMWDFGQSAPSELPKYYATID	648
Qy	767	VGRFWLGNK-LTLGGMARYEFGKSIRATAERYIDGTNGGNTSNVROLGKRSIKOFTLAR	825
Dd	768		
Dd	649	AGVLPDDRSLGTTTIOTGKRKASAGDFPD-----TGTIPMVQEKT-----	694
Qy	826	QPLIFDPAAYEPKKNLIFRAEYKNFLDRRYIDPLDAGNDAATQRYYSFDFPKDKDEEVT	885
Dd	827	:	
Dd	695	-PTVINAYANYQMNKVNLSKFAVHNLMMKNYSNALDRSNSAPLM-----DEQ----	740
Qy	886	CNADKTLNCNKYGGTSKSVLTJNFMARGTEFLTITSYKF	922
Dd	887	:	
Dd	741	-----GONKQT-----ARGRSFLFGGEIRF	760

RESULT 11

RESOLUT. 1.1  
US-09-252-991A-28261  
; Sequence 28261, Application US/09252991A  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; TITLE OF INVENTION: AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788

```

, PRIOR FILING DATE: 1998-02-18
, PRIOR APPLICATION NUMBER: US 60/094,190
, PRIOR FILING DATE: 1998-07-27
, NUMBER OF SEQ ID NOS: 33142
, SEQ ID NO 28261
, LENGTH: 977
, TYPE: PRT
, ORGANISM: Pseudomonas aeruginosa
US-09-352-991A-28261

```

Query Match	8.6%	Score 422;	DB 16;	Length 977;
Best Local Similarity	22.6%;	Pred. No. 3.3e-29;		
Matches 214;	Conservative 115;	Mismatches 382;	Indels 238;	Gaps 37;
Qy	30	RAGSEAOIQVLEHHVAKRVPKDKKFTDARAVST--RODIFKS-SENLDNIIVRSIPGA	86	
Db	211	RVSQDDLVQSPSV-ISAAR--PDWVYQTPHSVSVIGRQIERNPPRHAADMLEETPGV	267	
Qy	87	FTQODKSSGIVSLNIRGDSGFRVNTWVDGITQTFYSTSDAGRAGSSOFGA-SVDSNF	145	
Db	268	YSSVSQODPGLSVNIRGTDQYGRVMSVDGMRQNYQ-----QSGHQQRNGTLVDPPEL	320	
Qy	146	IAGLDVYVKGFSFGAGINSLAGSANLITGLVDVVQGNNTYG---LLLKGLTG-TNSTK-	200	
Db	321	LSEVVIDKGASSMGAGVIGGIANFTVRDLRPGKQVGGVRVITSLGGDANGTHF	380	
Qy	201	-GNAMAAIGARKWLSEASGVLYGHGSRSSVAQRYRVGGGQHIGNFGAEYLERKKORYF	259	
Db	381	IGSAFAIGTEVW-----DMLVAASERHLG-DYDPGTKGS-IG-----ELRTGAWF	424	
Qy	260	VOEGGLKFNSSNGKWERDFORPYWKTKWYQKYNDDPQELQYIEGHDKSWRENLAPOQYDIT	319	
Db	425	NPEAGORVK-----HSPVAYSQVY-----MSRLAKLGVAL	455	
Qy	320	PIDPSSLKQOSAGNLFKLEYDGVFNKYTAOFRLDNTKIGSRKIINRYQNYGLSL-NSY	378	
Db	456	PDQD---RLQFSYLTTOVSYYDDA-NMLNTENQALWEKLSGSDVRAQNAFDYGVAPDNL	511	
Qy	379	ANLNLTAAY--NSGRQYKPGSKFTGNGLKDFFTYNNAKILDINTATRLPRETELQT	436	
Db	512	VDFKAKLYVYDNRNRQOTLQRGITPGYSITYQTDTYG---AQANTSTFALDDLSTLRA	567	
Qy	437	TLGFNYFHEYGNKR-----PEELGLFFDQDODNGLYSLGRFKGDKGLLPQ	485	
Db	568	NYGLEFFYDKVRPDSOSPRASTSAVGPPAAEGM---TPKGRALGSLFARLDYD-----	618	
Qy	486	KSTTVQAGSQYFNFTFYDAAKKDIYRL-----NTSTNTGYRFGGEYTYGYSDDF	539	
Db	619	-----YDDMLNLNAGLRVDRYRLRGDTGFNARTFLIGTTROTDMPLQYAVD---	664	
Qy	540	KRAFGENSPYKKHCNOSGIIYEPVLKKYCKKRANNHUSVISADFG--DYFWPFPASYSRTH	598	
Db	665	-REBGRFSPT-----GLSVKPGVDWLQLEFATYTKGW	695	
Qy	599	RMPIQIEMYSQIGDSG-----VHTALKPERANTWQGFNTKKGLLKODDTLGLKLVG	652	
Db	696	RPPAVTESLITGRPHGGGAENMYNPFSLPERSKAWEGFNVULKENLWFSDRLGLKLVAY	755	
Qy	653	YRSRIDNYIHNVYKGWDLNGLNIPSWSSYSTGLAYTIOHRNFKDKVHKHGFLELNTYDGR	712	
Db	756	FDTRVDDFIEMGMGMQ-----PPGYGMAGIGNSAYVNNL-DSTRPRGVEYQLDYDAGL	807	
Qy	713	FTFNLSTAYOKS-----TQPTNFSDAESPNNASKEDOLKOGYGLSRV---755		
Db	808	AYGOLSYTHMIGSNDFCSTAWLGGVYQTVKSGRRPPVIDMRPDEQANAATHCSAVLGS	867	
Qy	756	SALPRDYGRLVETRWLGNKLTLGGAAMYRFGKSIATAEERYIDGTNGGNTSNVROLGK	814	
Db	868	AEHMPDRG-----SLTLG--MRFFDRDLVGARARYSEGYSVAGGATVSAGV	914	
Qy	815	RSIKQTETLARQPLIFDYAAYEPKKNLIFRAEYKNIIFDRRYIDPDLDAAGNDAATQRYSS	874	

Db 915 YPADWKEY-----TVYDLYGSYRVSDDELTLRLAMENVTDRAVLVPLG----- 956

QY 875 FDPKDDEEVTCTADKTLNCKYGGTSKSVLT - NFARGRTFLITMSYKF · 922

Db 957 -----DVLAFTLGRGRTLQGTLEYQF 977

RESULT 12

US-09-252-991A-27095

; Sequence 27095, Application US/09252991A

GENERAL INFORMATION:

; APPLICANT: Marc J. Rubenfield et al.

;	TITLE OF INVENTION:	NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS
:	TITLE OF INVENTION:	AERUGINOSA FOR DIAGNOSTICS AND THERAPEUTICS

FILE REFERENCE: 107196.136

; CURRENT APPLICATION NUMBER: US/09/252,991A

; CURRENT FILING DATE: 1999-02-18

; PRIOR APPLICATION NUMBER: US 60/074,788

; PRIOR FILING DATE: 1998-02-18

; PRIOR APPLICATION NUMBER: US 60/094,190

; PRIOR FILING DATE: 1998-07-27

; NUMBER OF SEQ :

; SEQ ID NO 2709

; LENGTH: 992

```

; TYPE: PRT

```

ORGANISM: *Pseudomonas aeruginosa*

US-09-252-991A-27095

Query Match	8.6%	Score 421;	DB 16;
Best Local Similarity	23.2%	Pred. No. 4.2e-29;	Length 992;
Matches 226; Conservative	129;	Mismatches 370;	
		Indels 250;	Gaps 42;

QY 47 AKRVPKDKKVFTDARAVSTRODIFKSSSENLD-----NIVRSIPGAFTOODKSSSGIVS 98

Db 169 AELADPQKETVT-----APRSSVYLSSSEDIIDFRGRVSUGLLOGIPGVGVDSRNGGALD 2233

QV 99 LNIRGDSGFGRVNTMVDGITOTFYSTSTDAGRAGGSSOFGASVDSNFTAGLDVVKGSESG 158

47 :||||| |||||  
55 :||||| |||||  
62 :||||| |||||  
Db 224 VNIRIGQGSRVAVRVDGAEQ-----ALDVRSGYAGTQRSYIDPDLVSSVTVDKGPSTR 278

QV 159 SAGINSLAGSANLRTLGVDVVOGNNTYGLLLKGLTGNTSTKGNAMAA TGARKWJ.FSGAS 218

QI	100	SAGINSTRASAMUKALISVDDVVQQGNNIGLLKRGGLIGINSRGNMAAALGAKRWDESSGS	
		: : : :	
D <sub>b</sub>	279	SGAI---GGSVEMRTGVKDILVDGKDLGVRETGDVNN-----GVAPHRSSASSKTENIUSS	332

OV 219 V-----GVIYGHSP--SVAQNYRVGGGOHTGNECAEVI.FRPKORVEVOEGCI.KENSNS 271

[illegible]

07 272 CKWERDEORBYWK-----TKWY-----OKYNDRORFIKUTETCHUSWBENI ARDYDIT 310

04 320 ----- DTDDST YAOOCNCN EWT EYDOW - ENKVTACDDNY NMVTCCKYTNWNY 367

QY	320	-----PIDPSSLKQQSAGNLFKLEYDGV-FNKYTAQFRLDNTKIGSRKRIINRY	36
Dh	433	IGYRVYDGRGTGRTIMPSDPIREFGTAGTVOYDLSNVKIDTWTATPVYVLPEN-----NDIV	485

Q: 268 QENVCT CT NEWAVANT NY 11753 VANCOCDOZVDVCCCKENMOLCOT Y KDPBETWANNXVXT T OF NINTERFED A 00

QY	308	QFNTGLNSNYANLENLTAYNSGRQYPRGSKFTGWGLKDFETYNNAKI-LDNNNTATE	428
Dh	486	DI SMICI LMMMDAYKSDMI MEVIT A DDCAAAYDCDNNM-----DQNDNPTCCCT NNTA DE	536

[illegible]

```

Qy 427 RL-----PRETELQTTLGFN-----YFHNEYGKN----REPEE 455
      ::::: | | | | |
Db 527 YMRGRGRGRKRIKYACCGRQVMDTQRGGKGMHMYNYNYMRYDRTSMTSGYAYWYK 506

```

QY 456 LGLFEDG-----PDQDNGLYSYLGRPKGDKGLLPQKSTIVQPA--CSQYF---NTFYFD 504

Qy 505 AA---LKKDIYRLNYSTNTVGYRFGCEYTGYYGSDDEFRAGFGENSPYKKHCHNQSCGIY 561

```
Db 216 -----DKTNR-----EPNSQLM 228
Qy 309 RENLAPQYDITPIDSSLKOOSAGNLFKLEYDGVF--NKYTAQPRDLNTKIGSRKIINRN 366
Db 229 KINKP-----NDFHELSGRFYHNKFT-----KRHDSYD 260
Qy 367 YQFNPGLS--LNSYANLNLTAAYNSGRQYKPKSGKFTGWLKDFETYNNAKILDLNNTAT 425
Db 261 YLKYHVTTPFSELIDTNILLGSGKGNQYFVKMSGLGK-----ESHKNSNTIDIKNTSR 315
Qy 426 FRLPRETELQTTLGFNFHNEYGK 449
Db 316 FNY-GETDISFTLGSKLMDEYHK 338

RESULT 14
US-60-215-161-4927
; Sequence 4927, Application US/60215161
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; APPLICANT: Hinkle, Gregory J.
; FILE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; CURRENT APPLICATION NUMBER: US/60/215,161
; CURRENT FILING DATE: 2000-06-30
; PRIOR APPLICATION NUMBER: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 4927
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-60-215-161-4927

Query Match 8.2%; Score 402; DB 26; Length 364;
Best Local Similarity 28.4%; Pred. No. 4.8e-28;
Matches 109; Conservative 57; Mismatches 126; Indels 92; Gaps 11;

Qy 74 ENLNVIRSPGATQODKSSGIVSLNIRGDSGFRVNTWVDGTTQTFYSTDAGRAGG 133
Db 39 ESMDSVLSRLPGTYTQMDTSQGTIAVIRGMSGFGRVNMVYDVSQSFYGIAPSEFAHT 98
Qy 134 S--SQFGASVDSNFIAGLDVVKGSFSGAGINSLAGSANLRTLGVDDVVOGNNYTGILLK 191
Db 99 QPYNQTAGLIDSNFIITDVRGQANDSDSVNALVGSANFTIGIDVIFEGNKLGLTLK 158
Qy 192 GLTGTNSTKGMAAIGAR--KWLESASVGLYGHRSRRSVAQNYR--VGGGGQHGIFGA 248
Db 159 SAYGTNGLKGNMTAIGRTQAFTGEGSICAMLAISHSDAHYKNAMGVSSSEFGT--- 215
Qy 249 EYLERRKQRYFVQEGGLKFNNSGKWERDFQRYWKTKWYQKYNDPQELQKYTEGHDKSW 308
Db 216 -----DKTNR-----EPNSQLM 228
Qy 309 RENLAPQYDITPIDSSLKOOSAGNLFKLEYDGVF--NKYTAQPRDLNTKIGSRKIINRN 366
Db 229 KINKP-----NDFHELSGRFYHNKFT-----KRHDSYD 260
Qy 367 YQFNPGLS--LNSYANLNLTAAYNSGRQYKPKSGKFTGWLKDFETYNNAKILDLNNTAT 425
Db 261 YLKYHVTTPFSELIDTNILLGSGKGNQYFVKMSGLGK-----ESHKNSNTIDIKNTSR 315
Qy 426 FRLPRETELQTTLGFNFHNEYGK 449
Db 316 FNY-GETDISFTLGSKLMDEYHK 338
```

RESULT 15

```
US-09-897-516-6415
; Sequence 6415, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; FILE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 6415
; LENGTH: 937
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-6415

Query Match 7.9%; Score 386; DB 22; Length 937;
Best Local Similarity 19.9%; Pred. No. 7.2e-26;
Matches 216; Conservative 144; Mismatches 385; Indels 338; Gaps 40;

Qy 27 DAGRAGSEAOI-QVLEDVHVKAKRVPKDKVFTDARAVSTRODIFKSSSELDNIVRSIPG 85
Db 6 DKDEAGYDAVYDKDISNIYIGKKEIER-----YKGASPAD-VIRGAVG 47
Qy 86 AFTQODKSSGIVSLNIRGDSGFRVNTWVDGTTQTFYSTDAGRAGGSQFGASVDSNF 145
Db 48 YSGDARNSCALDINIRKIOGQGRIPVTIDGTEQ-----AITVGRGYNANRRNYIDPNL 102
Qy 146 IAGLDVVKGSFSGSAGINSLAGSANLRTLGVDDVVOGNNYTGILLKGLTGTNSTK----- 200
Db 103 ISSIEIEKGPRLNRKVGSGVGAISIKTLNIDDVVPEGETFGINTKLETSSNSVKERTPS 162
Qy 201 -----GNAMAAIGARKWLESASVGLYGHRSRRSVAQNYRVG--GGGQHG 245
Db 163 LSLGQDYRDVNPFIHNGIESDPALKITPHSSKDNKLEFGKDNA-----FRVAVGTRQYFD 218
Qy 246 FGAEYLERRRKQRYFVQEGGLKFNNSGKWERDFQ-----RPYW----- 283
Db 219 LMLAYAYRHKGNFYFAGKGG--AHRYDAAATEADINLMLNPKTSLDPLPFAARIYRPGNEV 277
Qy 284 -----KTKWYQKYNDPQELQKYTEGHDKSW-----RENLAPOYDI 318
Db 278 PNTSSKMOSVLIKNTWH--FTDEQALQALAFNRTRMEFGDIMPSRLASVLAKENSVPQWPL 335
Qy 319 TPIDPSLSLKQOSAGNLFKLEYDGVFNKY-----TAQFRDLNLTGIG--SRKIINRYQ 368
Db 336 -----ANARQQAASLANK--WDSAAANPYIDFNMNLWTTRTISNTNTSGGYPGPVTRDYD 388
Qy 369 F--NYGLSLNSYANLNLTAAYNSGRQYKPKSGKFTGWS--LLKDFETYNNAKILDLNNTA 424
Db 389 WEGGKGRSTNIDGTLINTAVTNAQNRR-----WGVDISNKKFELTQNLDLTLMGNFQ 439
Qy 425 TFRLPRETEL-----QTTLGFNY----- 442
Db 440 RERLGSDDDISHIDNLYFFQSPARKQORQEIENAFNFDWRPTSWALGAGAKRVSYWSKD 499
Qy 443 -FNE--YGNKRPEELGLFF-----DQPD-----ODNGLYSYLGR----- 475
Db 500 DFLNERRIARDRNYKEPGEIIGYKMSYRTVTEDEANDKRNDEKYEYKSLSTKRELR 559
Qy 476 -----FKGDKGLLPQKSTIVQ-----PAGSOYF-----NTFFVDAALKK 509
Db 560 AKTKKIKDRNDRIQQKRWLVFEKFWKYNPCTGKLNKSDNPFYNGQLDMNEKVIDPISGR 619
```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:54:04 ; Search time 54.24 Seconds  
(without alignments)  
2276.355 Million cell updates/sec

Title: US-09-762-926-4

Perfect score: 4904

Sequence: 1 MRSSRLPAPICFYLGMVLY.....SVLTNFAFGRTFLITMSYKF 922

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 410475 seqs, 133914855 residues

Total number of hits satisfying chosen parameters: 410475

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Pending Patents AA.New.\*

- 1: /cgn2\_6/ptodata/1/paa/PCT\_NEW\_COMB.pap.\*
- 2: /cgn2\_6/ptodata/1/paa/US05\_NEW\_COMB.pap.\*
- 3: /cgn2\_6/ptodata/1/paa/US07\_NEW\_COMB.pap.\*
- 4: /cgn2\_6/ptodata/1/paa/US08\_NEW\_COMB.pap.\*
- 5: /cgn2\_6/ptodata/1/paa/US09\_NEW\_COMB.pap.\*
- 6: /cgn2\_6/ptodata/1/paa/US10\_NEW\_COMB.pap.\*
- 7: /cgn2\_6/ptodata/1/paa/US60\_NEW\_COMB.pap.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	279	5.7	917	5	US-09-584-501A-11
2	269.5	5.5	908	5	US-09-584-501A-12
3	239.5	4.9	934	5	US-09-540-209B-7679
4	203	4.1	564	5	US-09-545-199F-30
5	199.5	4.1	1083	5	US-09-540-209B-9521
6	195.5	4.0	925	5	US-09-540-209B-6249
7	189	3.9	967	5	US-09-545-199F-32
8	184.5	3.8	1094	5	US-09-540-209B-8695
9	180.5	3.7	707	5	US-09-540-209B-10058
10	179.5	3.7	659	5	US-09-545-199F-105
11	177	3.6	723	5	US-09-540-209B-6565
12	177	3.6	833	5	US-09-540-209B-7143
13	172	3.5	702	5	US-09-540-209B-8474
14	172	3.5	783	5	US-09-540-209B-9746
15	166.5	3.4	930	5	US-09-584-501A-2
16	160.5	3.3	3241	1	PCT-US01-13240-1
17	160	3.3	1178	5	US-09-540-209B-7922
18	159.5	3.3	1043	5	US-09-540-209B-9606
19	159	3.2	760	5	US-09-540-209B-6609
20	159	3.2	1152	5	US-09-540-209B-10027
21	157.5	3.2	722	5	US-09-540-209B-9749
22	156.5	3.2	947	5	US-09-540-209B-9753
23	156.5	3.2	1009	5	US-09-540-209B-10198
24	155	3.2	1102	5	US-09-540-209B-9537
25	154.5	3.2	699	5	US-09-540-209B-6132
26	154.5	3.2	1111	5	US-09-540-209B-9178

27	153.5	3.1	1008	5	US-09-540-209B-8074	Sequence 8074, Ap
28	153	3.1	1119	5	US-09-540-209B-9958	Sequence 9958, Ap
29	152.5	3.1	1120	5	US-09-540-209B-9538	Sequence 9538, Ap
30	152	3.1	927	1	PCT-US01-13240-3	Sequence 3, Appli
31	151.5	3.1	840	5	US-09-540-209B-10242	Sequence 10242, A
32	151.5	3.1	1087	5	US-09-540-209B-7962	Sequence 7962, Ap
33	146.5	3.0	1102	5	US-09-540-209B-7438	Sequence 7438, Ap
34	145.5	3.0	982	5	US-09-540-209B-7556	Sequence 7556, Ap
35	145	3.0	989	5	US-09-540-209B-7130	Sequence 7130, Ap
36	144	2.9	860	5	US-09-540-209B-9959	Sequence 9959, Ap
37	143.5	2.9	528	5	US-09-540-209B-7591	Sequence 7591, Ap
38	142.5	2.9	1066	5	US-09-540-209B-8040	Sequence 8040, Ap
39	142	2.9	808	5	US-09-540-209B-6468	Sequence 6468, Ap
40	141.5	2.9	1140	5	US-09-540-209B-9401	Sequence 9401, Ap
41	141	2.9	790	5	US-09-540-209B-6713	Sequence 6713, Ap
42	141	2.9	845	5	US-09-540-209B-9999	Sequence 9999, Ap
43	141	2.9	1118	5	US-09-540-209B-9536	Sequence 9536, Ap
44	140.5	2.9	713	5	US-09-540-209B-7523	Sequence 7523, Ap
45	140	2.9	767	5	US-09-882-227-78	Sequence 78, Appl

ALIGNMENTS

RESULT 1

US-09-584-501A-11

; Sequence 11, Application US/09584501A

; GENERAL INFORMATION:

; APPLICANT: Lo, Reggie Y.C.

; APPLICANT: Schryvers, Anthony B.

; APPLICANT: Potter, Andrew A.

; TITLE OF INVENTION: TRANSFERRIN BINDING PROTEINS OF

; FILE REFERENCE: PASTEURELLA HAEMOLYTICA AND VACCINES CONTAINING THE SAME

; CURRENT FILING DATE: 1996-11-29

; PRIOR APPLICATION NUMBER: 08/753,759

; PRIOR FILING DATE: 1996-11-29

; PRIOR APPLICATION NUMBER: CA 2,164,274

; PRIOR FILING DATE: 1995-12-01

; PRIOR APPLICATION NUMBER: 60/008,569

; PRIOR FILING DATE: 1995-12-01

; NUMBER OF SEQ ID NOS: 68

; SOFTWARE: FastSeq for Windows Version 4.0

; SEQ ID NO 11

; LENGTH: 917

; TYPE: PRT

; ORGANISM: Neisseria gonorrhoeae

US-09-584-501A-11

Query Match

Best Local Similarity 5.7%; Score 279; DB 5; Length 917;

Matches 224; Conservative 127; Mismatches 323; Indels 376; Gaps 58;

QY	5	FRLPICFYLGMVLYHHSVAEDAGRAGSEAOIQVLEDVHVYKAK--RVPDKKKVFTDARA	62
Db	7	FRNLILCSLMTAL---PAYAENV-QAG-QAQEKOLDTIQVAKKQKTRRNEVTGLGKL	61
QY	63	VSTRQDIFKSSG-NLDNIVRSIPG-AFTQDK--SSGIVSLNIRGDSGFRVNTWDGIT	118
Db	62	VKTADTILSKQVLDIRDLTRYDPCIAVVEQGRGASSG---YSIRG-WDKNRVSLTVGLA	117
QY	119	Q-TFYSTSTDAG--RAGGSQFGASVDSNFTAGLDVYKVGSGSAGINSLAGSNLRTLG	175
Db	118	QIQSYTAQAALGGTRTAGSSGAINIEYENVKAVEISKGSNVEQSGALAGSVAFOTKT	177
QY	176	VDDVQVQNNYGLLLK-GLTGTNS--TKGNAMAA-IGARKWL--ESGASVLYGHSSRS	229
Db	178	ADDVIGEGROWGIQSKTAYSGKNRGLTQSIALAGRIGGAELLIRTGHRHAGEIRAH----	233
QY	230	VAQNYRVGGGQHGIGTCAEYLERKQRYFVQEGGLFNSNSGKWERDFQRPYKTKWYQ	289

```
Db 234 -----EAGRGVQSFNRLAPVDDGSKYAYFIVEBECK-----NGHECKAMP----- 276
Qy 290 KYNDPQELQKIEGHDK-----SWRENAPQVDITPIDPSSLKQOS-----AGNLF--KLEY 339
Db 277 -----KQDVGDQKQVSTYRDYTCNPFLEA--DPLSYESRWLFRPGFRFENRKH 326
Qy 340 DGVFNKYTAQ---FRDLNWKIGSKRII-----NRYQFNYSGLSNLYANL 381
Db 327 IGGILERTQQTDFTRDMTPAFLTKAVFDANQKQAGSLRGNGHKYAGNH-----KYGGL 381
Qy 382 NLTRAAYNSGRQKYPKSGKFTGWGLLKDFEYNNAKILDNLNTATFRPRETELOTTIGFN 441
Db 382 -----FTSGENNAVPVGAIEY-GTGVFYD-----ETHTKSRYLE 413
Qy 442 YFNEYGKNRPEELGLFFD--GPDQDNGL-----YSVLGRFKG 478
Db 414 YVYNADKDTWADYARLSYDRQGLGLDNHFOQTHCSADGSDKYCRPSADKPFYSI---YKS 470
Qy 479 DKGLLPQKSTIVQAGSYQNTFYFDAALKDIYRLNYSNTVGY-RFGG-----EYTG 532
Db 471 DRVIYGESHKLQAAFKKSDPTAKI-----RHNLSVN-LGYDRFGSNLRHQDY--Y 518
Qy 533 YGSDDEPKRAFGENSPYKHKCNOSCGIYEPVLKYYKKRA-----NNHSVTSAD--- 583
Db 519 YOSAN---RAYSLKTPP-----QNNGKTSPPNGREKNPYWVSTGRNVV 559
Qy 584 -----FGD-----YEMPEA-----SYSRTH----- 598
Db 560 TRQICLFGNTYTDOCTPRSNGKSYAAVRDNVRLGWADVGAGLRDYDRSTHSDGSGVS 619
Qy 599 -----RMPNIQEMYFSQIGDSGVHTALKPERANT 627
Db 620 TGTHTLSWAGIVLKPADWLDLTYRTSTGFRLPSPFAEMYGWSGDKIKAVKIDPEKSFN 679
Qy 628 WQFG-----FNTYKGLLKQDDTLGLKLVGRSRIDN-----YIH- 662
Db 680 KEAGIVKFGDGNLEASFWNNAYRDLIVR-----GYEAQIKDGKEQVKGNPAYLNA 730
Qy 663 -----NVYK--W---WDLNGNIP-SWVSSTGLAYT-IQHRNFKDKVHKHGFLELN 707
Db 731 QSARITGINILGKIDWGVND---KLEPGWYST--FAYNRVRVDIKRADRTDIQSHL- 784
Qy 708 YDYGREFTNLAYVOKTQPTNFSDASESPNNASKEDOLKOGYGLSRVSAALPRDYGRLEY 767
Db 785 FD-----AIQPSRYV-----VSGYDQPEGKMGVNGMLT-----YSKAKE 819
Qy 768 GTRWLGKLTILGGAMRYFGKSIRATABERYIDGTNGNTSNVROLGKRSIKQETLARQP 827
Db 820 ITELGSRALING-----NSRNTKATARRTRPW 847
Qy 828 LIFDFAAYEPKKNLIFRAEVKNLFDRIYI 857
Db 848 YIVDVSGYTVVKKHFTLRAGYNNLLNRYV 877

RESULT 2
US-09-584-501A-12
; Sequence 12, Application US/09584501A
; GENERAL INFORMATION:
; APPLICANT: Lo, Reggie Y.C.
; APPLICANT: Schryvers, Anthony B.
; TITLE OF INVENTION: TRANSFERIN BINDING PROTEINS OF
; FILE OF INVENTION: PASTEURILLA HAEMOLYTICA AND VACCINES CONTAINING THE SAME
; FILE REFERENCE: A34763 021645.0106
; CURRENT APPLICATION NUMBER: US/09/584,501A
; CURRENT FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: 08/753,759
; PRIOR FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: CA 2,164,274
; PRIOR FILING DATE: 1995-12-01
; PRIOR APPLICATION NUMBER: 60/008,569
; PRIOR FILING DATE: 1995-12-01
```

```
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12
; LENGTH: 908
; TYPE: PRT
; ORGANISM: Neisseria meningitidis
US-09-584-501A-12
```

Query Match 5.5%; Score 269.5; DB 5; Length 908;

Best Local Similarity 21.0%; Pred. No. 1.1e-12;

Matches 216; Conservative 116; Mismatches 354; Indels 343; Gaps 49;

```
Qy 5 FRLPICFVILMGVNLVHHSAEDAGRAGSAQIOVLDDVHVAK--RVPDKKVFETDARA 62
Db 7 FRNLICLSMTAL---PYAENV--QAEQAEKQLODTIQVKKAKKQTRDRNEVTGLGKL 61
Qy 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFRGFRVMTVDGIT 118
Db 62 VKSSDTLSKEQVLNIRDLTRYDPGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGVS 117
Qy 119 Q-TFYSTSTAG--RAGSSQFAGASVDSNFIAGLDVYKGSFSGSAGINSLAGSANLRTLG 175
Db 118 QIOSYTAQAALGGTRTAGSGCAINEIEYENYKAVEISKGSNSSEYNGALAGSAVAFQTK 177
Qy 176 VDDVVOGNNTYGLLLK-GLTGTNSTKGNAMAAIGARKWLESASGVLYGHRSRSVAQNY 234
Db 178 AADIIGEGKQWGTQSKTAYSGKHQALQSLALAGR---SGGAELIITYTKRRGRETHAH 233
Qy 235 RVGGGGQHIGNFGAAYLERRKQ-----RYFVQV---GGLKFNSEN-----SGKWDERF 278
Db 234 KDAGK--VQSFNRLVLDEBKKEGGSQYRYFIVEECHNGYAAACKNKLKEDASVKDER-- 289
Qy 279 ORPYWTKWTKYKNDPOE-LQKYIEGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKL 337
Db 290 -----KTVSTQDYTGSNRLLANPLEYGSQWL--FRPGW-----HLDNR 326
Qy 338 EYDGVFNKYTAQFRDLNWKIGSKRIINRNYQFNYSGLSNLYANLNTAAYNSGRQKYPK- 396
Db 327 HYVGAVLERFQQTDFTRDMIVPAFYFSEDYVPGSLKGLGY-----SGDKAERL 376
Qy 397 -----GSKFTGWGLKDDFETYNNAKILDLNNTATFRPRETELOTTLGFNYFHNEYGNKR 451
Db 377 FVQEGESTLOGIG-----YGTGVFYD-----ERHTKNRYGVVEYVYHNADKDT 418
Qy 452 FPEELGLFFD--GPDQDNGL-----YSVLGRFGDKGKLLPKST 488
Db 419 WADYARLSYDRQGLDLDNRLQOQTHCSHDGSKNCRPDGNKPYSF---YKSDRMIEESRN 475
Qy 489 IVQAPAGSQYFNT-----FYFDAAL----- 507
Db 476 LFOAVFKKAFDTAKIRHNLSINLGYDRFKSQLSHSDYIYLNQAVQAYDLITPKKPPFNGS 535
Qy 508 KKDIYRLNYSNTVGV---YRFGGE-YT-----GYGSDDEPKRAFGENSPYK 552
Db 536 KDNPYRVISGKTTVNTSPICRFGNNTYDCTPRNIGNGYI-----AAVDNVLGR 587
Qy 553 HCNQSGCI-YEPVLKKGKRRANNHVS-----ISADFGDYFMPFASYSRTH-----R 599
Db 588 WADVAGAGIRYD-----YRSTHSEDKSVSTGTHRNLNWSNAGVVLKPFMTWDLTYRASTGFR 642
Qy 600 MPNIQEMYFSQIGDSGVHTALKPERANTWQFG-----FNTYKGLLKQDDT 645
Db 643 LPSFAEMYGWGRACESLTKLDLKEPKSFNREAGIVFKGDFGNLEASYFNNAIRDLI----- 697
Qy 646 LGLKLVGRSRIDN-----YTH-----NVYK--WVDLNGNIPSWVSSTGL 684
Db 698 ----AFGYETRTQNGQTSASGDGPGYRNAQARIAGINILGIDHWGVMGLPGLYST-L 752
Qy 685 AY-----TIQHRNFKDKVHKHGFLELNIDY--GREFTNLISAYOKSQPT 728
Db 753 AYNRKVKADIRADRTFVTYSYLFDAVQPSRYVLGICYDHPDGIWINTWFTYSK----- 807
```

Qy	26	EDAGRAGSEAIQVLEDVHVHAKVRPKDKVFVTDARAVSTRODIF---- <td>81</td>	81
Db	99	EDLGVIKMAADAMVLDNIIITSS--VAVSRKT--PVAVSTVDPVFTEELGTQEPPEVLK	154
Qy	82	SIPCAF--TOODKSSGIVSLNIRGDSFGFRVNT--MYDGTQTQTFYSTTDAGRAGSSQVR	138
Db	155	STPGIYATKGGGFGDSKVNIR--GFKTENSAMTINGVPM-----NDMEWG	198
Qy	139	ASVDSNFIAGLDVVKGSFSSGAGIN-----SLAGSANLRTLGVDDVVQGNNTYGLLLK	191
Db	199	GIYWSNN--AGLSDVTRSMQVORGLGASKVAAPSGVSGSINVTNITDANKGGFVSYGMDN	257
Qy	192	GL-----TGTNSTK--GNAMAAATGARKWLSEGSVGLVYCHSRRSVAQNYRVGGGQHIG	244
Db	258	GLNKILFKVSTGLTKSGWATILGGTKWAD-----GYIOG-----TWYEA-----	297
Qy	245	NFGAEYLERRKQRYFVOEGGLKNSGSKWERDQRPYMKTKWYQYNDP-----QBL	297
Db	298	-----YNWFVNI--TKRENDNHOLSFTAFAPAQWHN--QRGKDGLTIEGWQEV	342
Qy	298	QYITEGHDKSWRENLAPOYDITPIDPSLSUKQOSAGNLFKLEYDGVFNKYTAQFRDLNTKI	357
Db	343	AKNYMNGEKPYRN--PTYG-----FGLNGQRKSSAYN-----VYNK-----	377
Qy	358	GSRKLIINRYQFNVGLSLNSYANLNTAAYSN--GRQYKPGSKFTGWLK--DFETYNNA	415
Db	378	-POLSLNHLWQINEKSSLS-----TALYASIGR-----GYGSQGLTSADRSMWYGS	424
Qy	416	KILDLNNT-----ATERLPRETTELQTLGFN--YFHNEYG--KNRPEELGLGF	460
Db	425	NNGLNMTFRKADGTFPAYDIIYALNREASENGSVYMAWSKSNFHNWYGLLSTYTKFGDIF	484
Qy	461	DGPDQDNGLYSYLGRFKGDKGLELPQKSTIVOPAGSYOFNTFFYFDAALKKDIYRLNYSTNT	520
Db	485	D-----FVGGIDYRYKGTHTNELVDLYGGDFYVDDSSRKSIVLASNNWAAA	530
Qy	521	VGYRFGGEYTYGSDDEFKRAFGENSPYTKKHCNOSCGIYEPVLKKYGKKRANHHSVI	580

## RESULT

US-09-545-199F-30  
; Sequence 30, Application US/09545199F

```

: GENERAL INFORMATION:
: APPLICANT: Lowery E., David
: APPLICANT: Fuller E., Troy
: APPLICANT: Kennedy J., Michael
: TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
: FILE REFERENCE: 28341/6227.NCP
: CURRENT APPLICATION NUMBER: US/09/545,199F
: CURRENT FILING DATE: 2000-04-06
: PRIOR APPLICATION NUMBER: 60/153,453
: PRIOR FILING DATE: 1999-09-10
: PRIOR APPLICATION NUMBER: 60/128,689
: PRIOR FILING DATE: 1999-04-09
: NUMBER OF SEQ ID NOS: 165
: SOFTWARE: PatentIn Ver. 2.0
: SEQ ID NO 30
: LENGTH: 564
: TYPE: PRT
: ORGANISM: Pasteurella multocida
: US-09-545-199F-30

```

Query Match 4.1%; Score 203; DB 5; Length 564;  
Best Local Similarity 20.2%; Pred. No. 1e-07;  
Matches 132; Conservative 84; Mismatches 273; Indels 1

[illegible]

Db 200 -----LLFGSVRN--ASNYTRPKSKILFSKNKQSGLIKVMQITPEHLTLTSSVYGIH 252  
QY 262 EGGLKFNSSGKWERDQRPYWKWKQYKYNDOELQYIEGHDKSRENLAPOYDITPI 321  
Db 253 KG-----WEP-----WAAK-RDVMSRPTETETEKHHYGDVAMWKRKL 287  
QY 322 DPSSLKQOAGNLFKLEBYGVFNKYTAQFRLNLTGKSRKILNRNYQFNGLSLNSYANL 381  
Db 288 -----YRD-----ORDESLSKYRYLPENKKNINL 312  
QY 382 NLTAAYNSGROKYPKSGKFTG--WGLL--KDFETYNNAKILDLNNTATFRLPRETELOTT 437  
Db 313 SVQLSYSKTEONDRHREKVTSSFLGTGKNSWITSYL--TFDISNTSLNIGR-AEHELL 370  
QY 438 LGFNYFNEYKGRPELGLFDPDODNGLYSYLGFRFGDKLLPKQSTIVQAGSOY 497  
Db 371 FGLQWLKNN--RNTLMYHKG---GVKKADYNYGFQPY-----YMPSGRQY 411  
QY 498 FNTFYDAALKKDIYRLNYSNTVGYRFGGEYTYGYSDDDEFFKRAFGEN-SPTYKKHCNQ 556  
Db 412 TQAFYLODQIKWQNFLE---TGIRY-----DHINNIGOKNLAPRYN---DI 452  
QY 557 SCGIYEPVLKYGKRRANNIS--VSIADFGDYMPFASYSRTHRMPIQEMY----FSQI 611  
Db 453 SAG-----HDYSQNTYNGWSYILGLKYDYNHLYLSLTFNFSKTRAPVIDEQYETQY 506  
QY 612 GDSGVHTALPERANTWQGFNTYKKGKLLKODDPLGLKLYGYSRIDNITHNVYG 666  
Db 507 SVSATSINLEKEMINQTRVGGIITLNLHFOENDAFQFRTTYFYNRGKNEIFKTRG 561

RESULT 5  
US-09-540-209B-9521  
; Sequence 9521, Application us/09540209B  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR  
; FILE REFERENCE: 2709,1001-001  
; CURRENT APPLICATION NUMBER: US/09/540, 209B  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 10444  
; SEQ ID NO 9521  
; TYPE: PRT  
; ORGANISM: B.fragilis  
US-09-540-209B-9521

Query Match 4.1%; Score 199.5; DB 5; Length 1083;  
Best Local Similarity 20.4%; Pred. No. 5.1e-07;  
Matches 199; Conservative 132; Mismatches 367; Indels 277; Gaps 57;  
QY 4 SFLRKPICFYLGMVLMYHSHYAEADGRAGSEAOIOLV--EDVHVKAKEV---PKDKKV- 56  
Db 86 SLKVAPNSQELSVYI-----GKQOTIKVSESTYNNVILKEDAEVLDEVVVVGSGSQKKV 141  
QY 57 -----FTDARAVSTRODIEFKSSENLNIIVRSIPGAFQODKs-----SGIVSLNIRGDSG 106  
Db 142 VTGAVGMVSAEALR--PVANASQALQGVV---PGLNLTVGNNGGALDGTLLNNIRAGT 197  
QY 107 FGRVN-----TWVDGITQTFYSTD-----AGRAGSSQFGASVDSNFIAGLDVVKG 154  
Db 198 IGDGSGSPLVLIDGIEGLDNTVPNDIESVVLKDAASAIYGARASFGVI--LVTTKS 255  
QY 155 SFGSSAGINSLAGSANLR--TLGVDDVVOGNNYTGILLKGLTGTNSTKGN-----AMAA 206  
Db 256 GKSGKTNV--SYSGSARFSAIDGVPDIM---DSY--TFAQYFNRAASANKGGGDFAPAVMER 310  
QY 207 IGA-----RKWLESASVG---VLYGHSRRSVAQNYRVGGG---GOH 242  
Db 311 IKAYQETLKTATVDNAGAGIWOQWANANGDTDWEEFYDHWAPSQEHNLSINGTDTQY 370

QY 243 IGNGFAEYLERRKQRYFVQEGGLKFN-----SNSGKWBER-DFORPYW 283  
Db 371 L--ISSFLD---QKGLMRHGKDKFQRYTLNGKITTAFTWFKYTYSTKWTREDFERP-- 423  
QY 284 KTKWYQKYNDOELQYIEG---HDKSWRENLAPOYDIT--PIDPSSLKQOOSAGNLFKLE 338  
Db 424 -----SYLTGNFFHNLARKWPHVDPNGFPMDEGEVEQEMENGKQNSQ 468  
QY 339 YGVFNKYTAQFRLD--NTKIGSRKILNRNYQFNGLSLNSYANLNLTAAYNSGROKYPKG 397  
Db 469 KDFYTNLOQLVFFPIKNNKILNGLSGSVRTTYQYQHWELVPY-----AYNAVADPY--- 518  
QY 398 SKFTGMLLKDFETYNNAKILDLNNTATFRLPRETELOTTLGFNYFNEYKGRPELGL 457  
Db 519 --TYVWDMGYSVAAGSSRVNE-----YSWKENYTYTNIYSDFK 556  
QY 458 LFEDGPDQDNGLY--SYLGRFKGD---KGLLPQKSTIVQAGSOYFNTF-----YF 503  
Db 557 QF-----DNGHYFKVMAGFNAELYKTRNITAENKTLITP--GVPTINTATDDPOAYGGA 609  
QY 504 DAALKKDIYRLNYSNTVGYRFGGEYTYGYSDDDEF--KRAFGENSPTYKKHCNOSCGIY 561  
Db 610 DNSVAGFFARVNNYSKD---RYMFEANGRYDGSRRFVCKERWG--FFPSFSAGWNIA---R 662  
QY 562 EPLVKYKKGKRRANNHVSISADFG-----DYFMPFASYSRTHRMPIQEMYFSQIGD 613  
Db 663 EPEMESFAER--INMGSLKLRASGQLGNTNTDAWYFF-----YQTMP--VGSNRYGWLNVG 715  
QY 614 SGVHTALKP-----ERANTWQFG-----FNTYKKGKLLKQDDTLGLKLVGYSRIDN 659  
Db 716 ERPNYATNPGISVSKKTWETVETWVGLDWSFFNRLSG-----SFDY 758  
QY 660 YIHNVYKGMWDLNAGNIPSWSSYGLAVTIQHRNFKDKVHKHKGPELELN---DYGR 712  
Db 759 FVRYTY---DMIGPAPELSLLGT--SVPKINNSD--MKSYGFELEVNWRDRICEVSYGA 811  
QY 713 FFTNLSYAYQKSTQPTNFDASPPNNAKEDQLKQCYGLSRV-----SALPRDYG 763  
Db 812 KKV--LSDDQOKILR---YPNDSYDVGSYKGEHLNDIWLTTIGIAKSOEEMDAHLAKVD 867  
QY 764 RLEVGTIRW-LGNKL--TLGGAMRYFGKSIRATAEERVIDGTNG--GNTSNRYOLGKRS--I 817  
Db 868 QSSVGTNMGVGDIMYADLDG-----DGKISNGTNKLGDTGDIYIIGNSPRF 914  
QY 818 KQETELARQPLIEDF 832  
Db 915 KYGITLDAAMKGFDF 929

RESULT 6  
US-09-540-209B-6249  
; Sequence 6249, Application us/09540209B  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES  
; FILE REFERENCE: 2709,1001-001  
; CURRENT APPLICATION NUMBER: US/09/540, 209B  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 10444  
; SEQ ID NO 6249  
; LENGTH: 925  
; TYPE: PRT  
; ORGANISM: B.fragilis  
US-09-540-209B-6249

Query Match 4.0%; Score 195.5; DB 5; Length 925;  
Best Local Similarity 20.0%; Pred. No. 8.4e-07;  
Matches 156; Conservative 104; Mismatches 257; Indels 263; Gaps 41;  
QY 34 EAQIQVLEDVHVAKRVPKDKKVFETDARAVSTRO-----DIFKSSENLDNIIVRS 82

Db 236 EETHMLDEVITISGRIONVKSTOLGAETLRPTOLKNIPMALGEVDILK-----MVQA 288

Qy 83 IPGAFTQODKSSGIVSLNIRGDSGFRVNTMVDGITQTFYSTS-----T 126

Db 289 LPGVKTVEASSG---FNVRGATQDNLILLNDG---TIYNPHLFGFFAAFNDSMVKEA 342

Qy 127 DAGRAGSSOFGASVDSNF-IAGLDVVKGFSFGSAGINSLAGSANLRLTLGVDD-----VVQ 181

Db 343 EIKSSIPAGGRISILDITKEANKKEFTSAGIGLVTSKNLLEIPIIKORTSVLLS 402

Qy 182 GNNTYG-LLKGLTGTNSTK-GNA-----MAAIGARKWLES-GASVGVLYIGHRSRRAQNY 234

Db 403 GRTTYSWIMKQLPEKSEYKNGTAGFDVLAIAVAHFENDKHSNLVGYGYYSHDRFAFNSNE 462

Qy 235 RVGGGGHIGNFGAELERKQRYFVQEGGLKFNSNGKWERDQFRPYWKTKYQKYNBP 294

Db 463 KYG-----YNNLAA-----SARWRVFN-----480

Qy 295 QELQKYTE-CHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDL 353

Db 481 EKLIGFSAGYD-----HYD-----NNRETWNASTAYKLSFD--INQYF-----518

Qy 354 NTKIGSKIIINRYQFNYGLSLNSYANLNLTAAYNSGRQYKPKGSKFTGWLKLDFTYN 413

Db 519 -VKADFTNILLADKHTLNFPGKSMY-----HINSGTYE-PEGS-----ESFV 558

Qy 414 NAKILDNNT--ATFRLPRETELQTLGFNYFHNKYGNKRPPEELGLFPDGPDDNGLYS 471

Db 559 KDVLOKDALEAFYALGDWEITPKLSVN-----AGTY-----SLFS 597

Qy 472 YLG---REFGDKGLLPQKSTI---VQAGSOYFNTFYDAAKKDIYRLNYSNTVGYRF 525

Db 598 ALGPRSYQVAGMLPHESTITDTITAGAGKFMKTYH-----634

Qy 526 GGBY--TGYGSDDEKRAFGENS-----PTYK-----551

Db 635 GPEERLSARVAFYDNFSVAKAGFNSMRQYIHKLSNTVIMSPDTWKLSDVNIKPORGWQAA 694

Qy 552 --KHCNOCGCIYEPVLVKYKKGKRRNNHVSISADFGDYFMPFASYSKTHRMPNIQEMYFS 609

Db 695 AGLYLNPSPSGIWYSVBEYKFR-----MSDYLDY-RGGAKLLMHHIET--DVINT 742

Qy 610 QIGDSGVHTALKEP--RANTWQFGFNTYKKGKLLKQDDTLGLKLVGYSRSDNYIHNVYK 667

Db 743 QGHAYGVELQVKKOVGKLNGM--SYTYSRTFLRQND-----KRIEKPVN--GD 788

Qy 668 WDLNGNIPSWSSSTGLAYTIQHRNFKDKVHKHGFLELNYDYGREF--NLSYAYOKSTQ 726

Db 789 WYPTDYKPHDFKRVG-----NYK-FTHRYSMSINVDYSGRPTTIPAGQYIDESTQ 839

RESULT 7

US-09-545-199F-32

GENERAL INFORMATION:

APPLICANT: Lowery E., David

APPLICANT: Kennedy J., Troy

TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions

FILE REFERENCE: 28341/6227.NCP

CURRENT APPLICATION NUMBER: US/09/545,199F

CURRENT FILING DATE: 2000-04-06

PRIOR APPLICATION NUMBER: 60/153,453

PRIOR FILING DATE: 1999-09-10

PRIOR APPLICATION NUMBER: 60/128,689

PRIOR FILING DATE: 1999-04-09

NUMBER OF SEQ ID NOS: 165

SOFTWARE: PatentIn Ver. 2.0

SEQ ID NO 32

LENGTH: 967

TYPE: PRT

ORGANISM: Pasteurella multocida

US-09-545-199F-32

Query Match 3.9%; Score 189; DB 5; Length 967;

Best Local Similarity 19.2%; Pred. No. 2.9e-06;

Matches 210; Conservative 125; Mismatches 337; Indels 424; Gaps 54;

Qy 39 VLEDVHVHAKRVPKDKVFTDARAVSTRQDIFKSSE-NLDNIYRSIFGARTQODKSSGIV 97

Db 22 ILADSHQEAPEL-----DTITVSSQDDEMNIEKKIGETVKTASOLAKRQOVDSDRL 73

Qy 98 SLNIRGDSGFRVNTMVDGITQTFYSTSDAGRAGSSQFGASVDSNFIA-----GLDWW- 152

Db 74 --VRVETGV-----TVVEAGRFGSSGYAIRGVNDENRVAITVDGLHQAE 114

Qy 153 -----KGSFSSAGINSLAGSANLRLTLGVDDVVQGNNTYGLLLKGLTGTNSTKG----- 201

Db 115 TLSQOGKELFEYGNFNTRNSVEIETLKVAKIAKADSVKV-----SGSLGGAFLV 168

Qy 202 --NAMAAGARKW--LESASGVLYGHSRVSVAQNYRVGGGQHIGFGAEYLERK 255

Db 169 ETKDARDLFTEKDMHIGYKAGYSTADNOGLNAVTLAGRYQM-----FDALIMHSKR 219

Qy 256 QRYFVQEGGLKFNSNGKWERDQFRPYWKTK--WYQKYN-DPOELQKYI-----E 302

Db 220 HGHELENYDYKNGRDIQGEREKADPYTITKESTLVKFSPSTENHRFTVASDYLQHSR 279

Qy 303 GHDKSW-----RENLA-----POYDITPI-----321

Db 280 GHDLSYNLVATHTQLDEKESRHANDLTRKNVSFTYENTVTVPFMDTLKLSYSQORITT 339

Qy 322 -----DPSSLKQOS-----AGNLFKLEYDGVFNKYTAQFRDLN-- 354

Db 340 RARTEYDCGNELCDSYKNPLQFGKQGLDPAKNKIKQSGSL-----STQIVDENGKP 395

Qy 355 --TKIGSR-----KIINRYQFN-YGLSLNSY-----378

Db 396 FPPTTGTNNAAFSNNLRLRPTGFWLDCSVDFCNKPFVYINISNCTYQAREVILLSEETVD 455

Qy 379 ANNLTAAYNSGRQKY---PKGSKFTGWLKQDFTYNNNAKILDNLNTATFRLPRETELQ 435

Db 456 GKLYTAKESGGLPNYLILPNSKGYLPYD-YKERDLNTNTKQINLDTKTF-----505

Qy 436 TTLCGFNYFHN-EYG--KNRFEPEL--GLFEDG-----PDQDNGLYSLG--REFGDKG- 481

Db 506 --LTFNLENLSYGGVYSRIEKEKINKAGYEGRNPTWADRILGQSSYCGYNALCKPKHE 563

Qy 482 -----LLPQKSTIVQAGSOYFNTFYDAAKKDIYRLNYSNTVGYR-----FGGEYT- 530

Db 564 PLTSFLIPVEATT-----QSLYFANILKVH-----NMISIDLGYRDHIKYNPEYTP 610

Qy 531 -----GYI-----GSDDEFKRA-FGENSPYTKKHCNOCGCIYEPVLKKYK 569

Db 611 GVTPKIPDDMVKGLFIPMPKEPOLKDFYNVAKFGE--AYKK-----WKEYL 655

Qy 570 KRA-----NNHSVISADFG--DYFMPFASYSRTHRMPNIQEMYFS-QIGD 613

Db 656 PKNAEENIAYIAQDKTKFKHSYSLGATFDPLNLRVQVKYSKGFRAPTSDELYFTFKHPD 715

Qy 614 SGV--HTALKPERANTWQFGFNTFYDAAKKQDQDTLGLKLVGYSRIDNYIHNHYGKWWDL 671

Db 716 FTILPNPVLAPEEKAKNOEIALTVHDNMGFVSTSVFQTK---YRHFID-----759

Qy 672 NGNIPSWSSSTGLAYTI---QHRNFK-----DKVHKHGFLELNYDYGREF-----714

Db 760 -----LAYLSRNLNSVGGQAQARDQVYQNVNVDNAKVGLKLEINARLNLGYFHHVLDGF 815

Qy 715 -TNLSYAYOKSTOPTNFSADSESPNNA-----SKEDQLKQGYGLSRVS-----756

Db 816 NTSYKFTYQGRILD-----GDRPMNAIQPKASVFEGLGYDHKENKFGADLYITRVSEKKA 869

Qy 757 -----ALPRDYGRLEVGTWRWLGKNTLILGGAMRYFGKIRATAEERYIDGTNGTNSVRQ 811

Db 870 KDTYNMFYEQGYKDSAVRWRSDDYTLVDVAVGYI-----903

QY 812 LGRSIRKOTETLARQPLIFDYAAEPKKNLIFRAEYKNLFDRY-----IDPLDAG 863  
Db 904 -----KPIKNLTLOFGVYNLTDRKYLWESARSIKPRGTS 938  
QY 864 N-----DAATQRYYS 873  
Db 939 NLINOKTCAGINREYS 954  
RESULT 8  
US-09-540-209B-8695  
; Sequence 8695, Application us/09540209B  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR  
; FILE REFERENCE: 2709.1001-001  
; CURRENT APPLICATION NUMBER: US/09/540, 209B  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 10444  
; SEQ ID NO 8695  
; LENGTH: 1094  
; TYPE: PRT  
; ORGANISM: B.fragilis  
US-09-540-209B-8695

Query Match 3.8%; Score 184.5; DB 5; Length 1094;  
Best Local Similarity 20.1%; Pred. No. 8e-06;  
Matches 188; Conservative 132; Mismatches 312; Indels 303; Gaps 50;  
QY 5 FRKPCIFCYLMG-VMLYHSHVAEDAGRAGSEAOIQVLEDVHVAKRVPKDKVFTDARAV 63  
Db 9 YRLKSIFILFSCIFLOTAFAQONNVKITIKKNITLQEALREVERKQSDYLIAFNESKLE 68  
QY 64 ST-RQDIFKSSENDLVRSIPGAF-----TQDQSSGIVSLNIR 102  
Db 69 KTRVNLNINAESLDKTLASITGLSYKIDKIMIPQSKVEVESKRLSGIVK-DDK 127  
QY 103 GDSGFRVN-----TWVDG-----TQTFYSTSTSDAGRAGSSQF--- 137  
Db 128 GDPLIG-VNVSFKGSPGTGTGLDGRFSILAAKGNIEFVGTQYIIVGDASSLTVV 186  
QY 138 ---GASVDNFITAGLDVVRKSGSAGINSLAGSANLRTLGVDDVVGQNNITYGLLLK-- 191  
Db 187 LEEDAKALDEVVYVTLGIGKRAEKALSYVQVKSDA-----INDVKDANFVNGLTCKVA 240  
QY 192 GLTGNTSTKGNAMAAIGARKWLESAGS--VG---VLY-----GHSRKSVAQNVKVG 238  
Db 241 GVSINRSSGIG---GATRVVMRGAKSIVGNVNNVLYVDCMPTGPNPSKEINNDSYTPG 296  
QY 239 GGOHIGFAGYLE-----RRQRYFVOEGGLKFNSSNGK 273  
Db 297 GGEGISDFNEDIESILGPPAAALYGSAAANGVILINTKKG---OEGKLKISIN-- 351  
QY 274 WERDFQRYWKTWKYQKYNQDQELQKYEIEGDKSWRENLAPOYDITPIDPSSLQ-----Q 329  
Db 352 -NTEFTMPYVMEFQNYGN-----AKGYSKSWGEMLO-----QPSTFPRKDFK 395  
QY 330 SAGNLFKLEYDGVFNKYTAQFDRLNTKIGSRKII--NRNQFYNYGL-----SLNSVANLN 382  
Db 396 TGANIMNAANFSVGNKNQTFVSVAT--TNSTGIIPNNEYRYNFTLRNTASMLNDKLHLD 454  
QY 383 LTAAY-----NSGR-----QKYPKSGKFTGWLGLKDFEYNNNAKILDLNNTAT 425  
Db 455 LGASYVLQGDONMLSAGRYFNPLVLYLPRGEDFEA---VKYERY-----DTRN--- 502  
QY 426 FRUPRETELQTLGTFNFFHNEY-----GKNRFPPEELGLFFQPDODNGLYSYL-- 473  
Db 503 -KPIQEWISYCGDOGLN--LENPYVIVNREMFVSKKRY-----MFYANVKYD--ILSWLNI 553  
QY 474 -GRFKGDKGLLPK-----STIVOPAGS-----QVNTFYFDDAAL--KKDIYELNY 516

Db 554 AGRIRDVNTWTTSERKLNHASTIKLHAQSDKGAYNRSMEEQYQTYADIMLVNKNFGFNL 613  
QY 517 STN-----TVGRFGGE-----YGYGSDDEPKRAFGENSPYTKKHCNQSCGI 560  
Db 614 TANAGFSYEDHLTGTGMGIGGKLTVPNLFSAI-----NFDPASGPGSQSHTHTRNNSVFV 668  
QY 561 -----YEPVLKYGKKRANNSHSISADFGDYEMPFPASYSRTHRMPNIQEMY----- 607  
Db 669 STELGYSKMLYLTLTGRQEWASQLVNSDQTYFYPSVGVSGV-----ISEMVSILPKEISF 723  
QY 608 -----FSQIGDSGVHTALKP-----ERANTWQFGFNTYK 636  
Db 724 WKMRASFAEVGGPINVTGLTPTGVTDPKMGVINPISVYPPNFKABQTKSYELGTN--- 780  
QY 637 KGLLKQDDTLGLKLVGRSRIDN--YIHNVYGKWDLNGNIPSWVSSTGLA-YTIQHRNF 693  
Db 781 -----LRLFNKNKINIDATVYLTDTYNQTF-----LSSMSPASGYSGFYVQ---- 820  
QY 694 KDKVHKHGFLELNYD--YGR--FFTNLASYOKS 724  
Db 821 AGVRNKGIELSLGYNDRFGKGVGYATNLITYTANRN 855  
RESULT 9  
US-09-540-209B-10058  
; Sequence 10058, Application US/09540209B  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES  
; FILE REFERENCE: 2709.1001-001  
; CURRENT APPLICATION NUMBER: US/09/540, 209B  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 10444  
; SEQ ID NO 10058  
; LENGTH: 707  
; TYPE: PRT  
; ORGANISM: B.fragilis  
US-09-540-209B-10058  
Query Match 3.7%; Score 180.5; DB 5; Length 707;  
Best Local Similarity 21.0%; Pred. No. 8.9e-06;  
Matches 175; Conservative 101; Mismatches 332; Indels 225; Gaps 48;  
QY 36 QIOVLEDVHVAKRVPKDKKVFVDARAVSTRQDIEKSSENDLVRSIPCAFTQODKSSG 95  
Db 1 EVOI-----VSTRATSKTPVAFNVSKUELKONF--GQDIPFLLSWTPSALTTSAGAG 53  
QY 96 I--VSLNIRGDSGFRVNTWVGITOT-----FYSTSTD-----AGRA-G 132  
Db 54 IGYTTLVRGTDG-TRINITANGIPMNDABESHTLFWNNMPDFASSVKDIOVQRCAGTSTN 112  
QY 133 GSSQFGASVD-----SNFIAGLDVYKFGSGSAGINSLAGSANLRTL 174  
Db 113 GAGAFGASVNMQTEGISMQPYABINASYDSFNAHKETVK--FGTGLLKDHAFDARLSTI 170  
QY 175 GYDDVVQ--GNNTYGLLLKLTCTNSTKGNAMAAIGARKWLESASGVLYGHSRRSVAQ 232  
Db 171 GTDGYIDRASVDLYSYAQGGYFADNTSVKFTTFGKREKTYHA-----WNATKEEMK 223  
QY 233 NY--RVGGGGGQHLGNFG-ABYLERRKQRYFVQEGGLKFNEN-SGKWERDFQRPYWK--TKW 287  
Db 224 KYGPRENSCMYTDHGHITFYKQDQTDNYLQMNYYLLNHTFSAANNLNAALHYTKGDY 283  
QY 288 YQKYNPDQELQKYEIEGDKSWRENLAPOYDITPIDPSSL-KOOSAGNFKLYDGVFN- 344  
Db 284 YQYKEDRSLSKEY-----RLHPFMYDVGKEVEKSDLIROKKMDNHFQ---GGVFSV 330  
QY 345 KYTAQAFDLNTKIGSRKIINRNQFYNYGLSLNSYANLNLTAAVNSGRQKYPKSKETGWG 404  
Db 331 NYPHQNDAS-----LVNA-----LNYDGHGFRGVIVWK--NYIG-E 365

QY	405	LLKDFETY--NNAKILD--LNNATFRLPRTELETQTTLGPNFYHNE--YGKNRPFPEELGLFF	460
Db	366	LLPDHEYRNRAKKTDGNLYLKANYNLVAGLNAYADLQRYINYKTGH	413
QY	461	DGPDQDNGLSYLGRFGDKGLLPQKSTIVOPAGSOYFNFTFYFDAALKKDILYNLTNT	520
Db	414	-----DNKKDY-----NTDGL--QKLAV-----NDHFNFNPAGLNWDI-----DSNR	452
QY	521	VGYRFGGEYTGYYGSDDDEFKRAFGENSPTYKKHCNSCGIYEPVLKKYGGKRANNH--SVS	579
Db	453	V-----YASF-----SYAKEPTNNNYTDCNADEYPKAEKLYDYELCYTYRNTW	496
QY	580	ISADFGDYFMPPA--SYSRTHRMPINQMFPSQTGDS---GVHTAL--KPERANTWOQGFN	633
Db	497	LSAGVNFPYMDYLDQLVLITGELNEIGEAMARNVPDSYRTGVLMLGVKPCRWQWDI----	553
QY	634	TYKKGLLKQDDTLGLKLGVYSRIDNIHNHYCKWD-----LNGNIPSWVSSTGLAYT-	687
Db	554	-----NGTUS-----KNRVKNFTEKLYEDEVKNPIEVHGNTP-----IAFSP	591
QY	688	--IQHRNFKDKVHKHGFELNDY-----GRFTNLISYAYOKSTQPT	728
Db	592	DFILNRFSS-HK-GFEAAQLSQYSVKQYMSNAKOAEOQLDAYFVSNLNLAYTFOLRHV	649
QY	729	NFSDASESPNNAKEPOLKOGY---GLSRVSALPRDYG-----RLEYGTWLGN	774
Db	650	KSVTVGFTIYNLFNEKYENNGYAGSGVTCLKDGKPERYNYAGYAQAQNVTMGN	702
RESULT 10			
US-09-545-199f-105			
; Sequence 105, Application US/09545199f			
; GENERAL INFORMATION:			
; APPLICANT: Lowery E., David			
; APPLICANT: Fuller E., Troy			
; APPLICANT: Kennedy J., Michael			
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions			
; FILE REFERENCE: 28341/6227.NCP			
; CURRENT APPLICATION NUMBER: US/09/545,199F			
; CURRENT FILING DATE: 2000-04-06			
; PRIOR APPLICATION NUMBER: 60/153,453			
; PRIOR FILING DATE: 1999-09-10			
; PRIOR APPLICATION NUMBER: 60/128,689			
; PRIOR FILING DATE: 1999-04-09			
; NUMBER OF SEQ ID NOS: 165			
; SOFTWARE: PatentIn Ver. 2.0			
; SEQ ID NO 105			
; LENGTH: 669			
; TYPE: PRN			
; ORGANISM: Pasteurella multocida			
US-09-545-199f-105			

Query Match	3.78;	Score 179.5;	DB 5;	Length 669;
Best Local Similarity	20.1#;	Pred. No. 9.9e-06;		
Matches 168;	Conservative 96;	Mismatches 301;	Indels 269;	Gaps 38;

  

QY	109	RVNTMVDGIIQTFTSTDA	---	GRAGGSGQFASVDNSFIAGLDVVYKGSFSGSAGINS	164
		:	:	:	
Db	8	RVAVIDGIPQAESTISARYSTERHNGNIN	---	NTEYENVSSLKVQKGAASVWYGSGA	64
		:	:	:	
QY	165	LAGSANLRTLCVDDVGQNTTYGLLLKGLTGTNTKGNMAAIG	---	-----AR	210
		:	:	:	
Db	65	LGGTVEFTTKDIEDFVPRHGLFLSK	-	TGTYSKNREYROYVIGVGKGHEFFGFVQLTK	122
		:	:	:	
QY	211	KW----	LESGASVGLYCHSRRSQAQNY	-----RVGGGGQHICFGAEYLERKKORYF	259
		:	:	:	
Db	123	RWGHTINNGKTDLGEHRCKPNPLNYTTSWLTKVGYDINNTHRF	-	TLFEDREKKL	181
		:	:	:	
QY	260	VOEGLLFNSNGKWERDFQRPYWKTKWQKYNDPQELQKYGIEGHDKSWREN	---	LAPQ	315
		:	:	:	
Db	182	TEERTLGI-DAVAFAND-QTPYLRYGIEYRN	-----GLSWLETVKFLAKQ	227	
		:	:	:	

```

QY 316 YDITDIPSSSLKQOAGNLFKLEYDGVFNKYTAQFRLNLTIGSKRKLIINRNQFNVGLSL 375
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 -----KIQRSALQEF-----DIN-----NRN-----KL 246

QY 376 NSYANILNLTAAVNSGRQYPKSGKTGWG-----LLKDFETYNNAKILDNNATFRLPR 430
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 247 DSTMGFVILQRONIARGETSTPLY--WQPSHRHLSAKFE--FDKLENNKNIKFTFR--PW 302

QY 431 ETELOTTLLGFNFHNEYGNKRPEELGLGFDDGPDQNGLSYVLGRFGKDGKLLPQKSTIV 490
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 303 Q-----INFRQO--GRNNVTE-----V 318

QY 491 QPAGSOYFNFTFFDAALKDIIYLNYSTVTGYRFGGEYTYGYGSDDEPKRAFGENSPY 550
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 319 FVPKSRFESFLMDDIKIGELHLG-----LGRWDHY-----NY 353

QY 551 KKHQNQCIGYEPVLKKYKGRANHSVISADFGDYFMPFASY--SRTHRMNIOEMYF 608
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 354 KPLNLSQHNINQRLPY--PKTSSKFSYQLSLEYQLHPHQIAYRLSTGFRVPRVEDLYF 412

QY 609 SOIGDSG-----VHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLKLVGYRSRID 658
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 413 EDGRKSSSOFLNPDLQPETALNHEISYRFQNGYAHFSVGL-----FRTRYH 459

QY 659 NYIHNVYKQWDLNGNIPSWVSSTGLAYTIOHRNFK--DKVHKHGFLELNYDGR----- 712
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 460 NF-----IQREMTCDKI-----PYENRTYGYCTHNT 487

QY 713 --FFTNLAYOKSQPTNFDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGR 770
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 488 YVMFVNEPEAVIKGVEVSGALNGSAFGLSDGLTFLKGSYSKGQ-----NHDGDLKSIQ 543

QY 771 WL-----GNKLTLCGANRY--FGKSIRATAERYIDGTNGCGTNSVROLGKRSIKQ 819
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 544 WTVVTGIDYETEGSVSLSG--RYSAAKKAKDAIETETH-----DKKVVQK 588

QY 820 TETLARQPLIFDYAAEYKPKNLIFRAEYKKNLFDRRYIDPLDAGNDAATQRYYS 873
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 589 WPHLSPSYFVDETGOVNLKSNVILNMGVFNLENRDYM--TWDSAYNLFTRGYTS 641

RESULT 11
US-09-540-209B-6565
; Sequence 6565, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 6565
; LENGTH: 723
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-6565

```

	Query Match	3.6%	Score 177;	DB 5;	Length 723;
	Best Local Similarity	19.9%;	Pred. No. 1.7e-05;		
	Matches 181;	Conservative 121;	Mismatches 317;	Indels 290;	Gaps 48;
QY	22 HSYAEDAGRAGEAQIOVLEDV-HVKAKR-----VPKD--KKYFTDARAVSTQRDIK 71				
	:   :   :   :   :   :   :   :   :				
Db	24 HAQEPDSLKAVALSEVVVETSYOHLKNKNSTWRMEVGKEFLREHFTG----- 71				
	:   :   :   :   :   :   :   :   :				
QY	72 SSENLDNIVRSIPGAFTQQDKSSGISVLNI RGDSGFGRVNTMVDGITOTFYSTSDAGRA 131				
	: :    : :    : :    : :    : :    :				
Db	72 ---NLIQTLGTLPGVHS-MDIGSGSKPMIR-GMFNRISVVENGIKO-----EQQQ 118				
QY	132 GGSSQFGASVDNSNFITAG-LDVYKGFSFGSAGINSLAGSANRLTGLVDDVVGQNTTYG--- 187				

Db 119 WGADH-GLELDA-FNAGQVSRKGFASLLYSGDAMGGAIELVPL---PLPAGNRLFGEAS 173  
Qy 188 LLLKGLTGTSTKGNAMAAIGARWLESASGVLYGHSRVSVAQVYRVGGGQHIGNFG 247  
Db 174 LLGKSVNGT--LGSMLGKIKDAM-----YTWARYS-----EHOFGDYR 211  
Qy 248 AE-----YLERKORYFVQEGGLKFNNSGKWERD-----FORPYWKTWKYQY 291  
Db 212 IPTDTIVLTQMPVYHRR--LK--NTAGFERDVSWAAGFRKERYVSSVWVSNVQKT 265  
Qy 292 N-----DPQELQYIEGHKSWRENAPQYDITPIDPSSLUKQOSAGNLFKLEYDGV 342  
Db 266 GFFPCAHCIPDVSRLQ-----DDGDSRNIELPYSQVNHILKVSQTRSLLYDKWALTWDIG 319  
Qy 343 FNK-YTAQFRDLNLIKSRKIINRYQFNGLSLNSVANLNLTAAYNSGRQKYPKGSKFT 401  
Db 320 FOKNREWSRFHYDAQVPDKDPKDELAFNTLNTSSAVKLKLFASVWQHT-----A 374  
Qy 402 GWGLLKDFETYNNAKILDLNNTATFRLPRETELOTTLGFNYFHNEYGNRFPPEELGLFFD 461  
Db 375 GWDV-----QYORNTIAGYSFLLPAY--RRF--TTGAFW- 404  
Qy 462 GPDQDNGLYSLGRFKGDKLLPKOKSTIVQAGSQYENTFFDAAKKDYRLNYSNTV 521  
Db 405 -----MTYRPG-----PTLSFGGLRYDYCKIDASAYTD 434  
Qy 522 GYRFGGEVGYGSDDEFKRAFGENSPTYKKHCHQSCG-----IYEPVLKKYKKRANHS 577  
Db 435 PYL--AIYLRQGVGDEFIRKYEWSYVVRHRHFGDYSGSLGWSP-----SGHL 483  
Qy 578 VSISADFGDYMPPFASYSRTHRMPIQEMVFSQIGDSGVH-----TALKPERANT 627  
Db 484 LQVNV-----GHSFRLPGANE-----LASGVHHTGPRHEQGDAALASERG-- 524  
Qy 628 WQGFEN-TYKGLLKQDDTLGLKLVGRSRIDNYIH-NVYKKWDL--NGNIPSWVSGTG 683  
Db 525 WQFDASYTYENG-----PLSVLSLSPVSWFSNYIFLRPTGEWSILPHAGQIYRYTGAEA 578  
Qy 684 LAYTIQHRNFKDKVHKHGFEELENDYGRFF--TNLSYAYQKSTQ--PTNFSDAESP 737  
Db 579 LF-----AGGEAAVGDIFRHFNYRVSGEYVYTNCDHEIPLSFPSPASLR 624  
Qy 738 NNASKEDQLKOGYGLSRVSALPRDYGRLEVGTWRWLNKLTGGAMRYFGKSIRATAERY 797  
Db 625 NTLT-----WRYKEFSYIGEVQVHIAAQHRVARNEDP 655  
Qy 798 IDGT---NGGNTSVNRLGKRSIKQETTLARQPL-----IFDFYAA-----EPKKN--L 842  
Db 656 TPGAQLLNAGVSANLRIGG---IWAEVTLARNLSGAKYFNHLSFYRKVEIPEPGRNQI 712  
Qy 843 IFRAEVKNL 851  
Db 713 LIKVPFKSL 721

## RESULT 12

US-09-540-209B-7143  
; Sequence 7143, Application US/09540209B  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Bretton  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FR  
; FILE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS  
; FILE REFERENCE: 2709.1001-001  
; CURRENT APPLICATION NUMBER: US/09/540, 209B  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 1044  
; SEQ ID NO 7143  
; LENGTH: 833  
; TYPE: PRT  
; ORGANISM: B. fragilis  
US-09-540-209B-7143

Query Match 3.6%; Score 177; DB 5; Length 833;  
Best Local Similarity 19.4%; Pred. No. 2.le-05;  
Matches 184; Conservative 119; Mismatches 349; Indels 298; Gaps 42;

Qy 9 PICPYLMGVMLYHHSYAEDAG-----RAGSEAQIQVLEDVHVRAKRV-----K 52  
Db 51 PVAIAIEGV--YIGGYTNENGVIHNDVPTGSGTIVVSGIGVTKKVPPIHVTAGKVNRP 108  
Qy 53 DKVFTDARAVSTPQDIFKSE-----NLDNIVRSIPGAFTOODKSSGIVSLNIR 102  
Db 109 DIEDTQAELEEVQVIGKSEARRQEQAYAISVDLTKKAYNSAAPLNKLLNNVSSVRI 168  
Qy 103 GDGSGF-----RVNTMVDGITTQTFYSTSTDRAGSGSSQFGASVDSNFTAGL 149  
Db 169 EEGGMSGNFSLNGFSGNQVKFELDIPMDNF-----GSSFNLANISANMAERV 218  
Qy 150 DVYKGFSGSAGINSLAGSANLRTGLGVDDVVQGNNTVGLLKLGLTGNSTKGNMAAIGA 209  
Db 219 EVYKGLVPLNGLADALGGAVNIVRRDANLDATYSFG-----SFTHKVSVNGA 268  
Qy 210 RKWLESCASV--GVLYGHSRVSVAQVYRVGGGQHIGNFGAEYLERKORYFVQEGGLKF 267  
Db 269 YTHLKTGFTVRANAFNYS-----DNDYKV-----FVPIIDLAT 302  
Qy 268 NSNSGKWERDFORPYWKTWKYQYNDPQELQKYIEG-----HDKSWRENLAPOYDIT 319  
Db 303 NKKIDE-----RWVRRFND-----AYRSGGIRLETGITNKPADYLLAGILS 345  
Qy 320 PIDPSSLUKQOSAGNLFKLEYDGVFNKYTAQFRDLNLTIGSKRKII-----NRNQFNVGLS 374  
Db 346 KND-----KDVQTGATMDAVYGGV-----KMKSESVIPSYRYKKDGLFLDGLS 388  
Qy 375 LNSYANLNLTAAYNSGRQKYPKGSKFTGWLKDFETYNNAKILDLNNTA--TERLPRETE 433  
Db 389 LSLYG-----TNSVNTFVNDTIARRNWLGESV 417  
Qy 434 LQTLGFNYPHENYGNRFPPEELGLFDPDQDNGLYSYLGRFKGDKGLLPQKSTIVQPA 493  
Db 418 PPSISAGEYITDSKIKNR--EWLG-----NGNISYV-----IDGHSLLNHHV 458  
Qy 494 GSQYNTFFYPDAALKKDIYRLNYSNTVGYRFGGEYTYGYSDDDEFKR-----AFGE----- 545  
Db 459 VSMARRTM-----NDKVRPDDENNPNVPOQLTKNITG--LGWQIRYDRWANVFGKMYKL 510  
Qy 546 NSPTYKKHCHQSCGIYEPVLKKY-----CKKRANHSVSIADFGDYFMPFA-----SYS 595  
Db 511 YSITYKR-----LDEYTENARWEKVRDHKTNFGYGAAATYIILPSLOAKFSYE 558  
Qy 596 RTHRMPIQEMVFSQIGDSGV---HTALKPERANTWQGFN---TYKKGLLKQDDTLGLK 649  
Db 559 HAYRLPESIEMF---GDGLIQORNPDLKPSSRNLLNGLSFQTFGAHQLSADGNF--- 611  
Qy 650 LVGYRSRIDNYIHNYGKWWDLNAGNIPSVSVSTGLAYTIOHRNPKDKVKHKGHELELNVD 709  
Db 612 --IYRYTTDFILKGV-----SLTSNPTTGYENLG-----KVLTKGVEAAVRYN 652  
Qy 710 Y-GREFTNLSYAYQKSTOPTNFSDASES-----PNNAKEDQLKOGYGLSRVSALPRD 761  
Db 653 YKOLFHTGAGFTYQDITDRORYEKTDFSVEGEGITENITYKE-----RLNPILPYL 702  
Qy 762 YGRLEVGT---WLGKNTLTGAMRYF-----GKSIRATAERYA--DGTNGG 804  
Db 703 FANADAGVRPHDLIWRNSVLTFDYNLNYIHSYVLSFPLGAKSKSKVIPQFSDHALGY 762  
Qy 805 NTSNVRQLKRSIKQETTLARQPLIFDFYAAYPEKKNLIFRAEVKNLFD 854  
Db 763 SMDN-----KYSYVVECTNLNOKLYDNRLQKPGR--AFNVKLRYFFSK 806

## RESULT 13

US-09-540-209B-8474  
; Sequence 8474, Application US/09540209B



## ; GENERAL INFORMATION:

; APPLICANT: Gary L. Breton  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FRA  
; FILE REFERENCE: 2709.1001-001  
; CURRENT APPLICATION NUMBER: US/09/540,209B  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 10444  
; SEQ ID NO 8474  
; LENGTH: 702  
; TYPE: PRT  
; ORGANISM: B.fragilis  
US-09-540-209B-8474

Query Match 3.5%; Score 172; DB 5; Length 702;  
Best Local Similarity 18.9%; Pred. No. 4.2e-05;  
Matches 143; Conservative 107; Mismatches 267; Indels 240; Gaps 36;

Qy 170 NLRTLGVDVVGGNNYGLLLKGLTGNSTKGNMAAIGARKWLESASGVLYGHRRS 229  
Db 95 SLNLYQSDAVK--HFAGVTYKDYGGIGGLKTVSLRSLGAEH-----TAVG--YDGITIS 145  
Qy 230 VAQNYRVGGGQHIGNGAEVLER-----RKQRYFVQEGGLKFNNSGKWERD 277  
Db 146 DCQT-----GQIDIGRSLQNDVRLSLNSGSDNIFQPARFFASAGILNQTLT----- 194  
Qy 278 FQRYPKTKYQKYNPQELQKYIEGHDKSWENLAPQYDITP-----IDPSSLKQOQSAGNL 334  
Db 195 ---PQFK-----DNRRTNLSASFSGTSGVLYNPSSLLEQLSRK 230  
Qy 335 FKLEVDG-----VFNKYTAQFRLNTKIGSRKLIINRYQNYGL--SLNSYANLNL 383  
Db 231 WLSANGEMWSADGHYFTLHYGDNLDLTGREKKNTEVKNLRAEAGLFGNFSDEQWRL 290  
Qy 384 TAAVNSGRQKPKGSKTGWLLKDFETYNNAKILDLN-----NTATFRLPRETELOTLG 439  
Db 291 KAYYQSSRGLPNATY-----YDYSQHLWDKNVQVQSKFEFSQWVQISAK 342  
Qy 440 FNYPHNEYGNRPFPEELGFLFDGPDQDNGLYSYLGRFKGDKGLLPQKSTIVQAGSQYFN 499  
Db 343 WWSYQRY-----LD-PD-----YKSGEG-----KTENSYYQQ 369  
Qy 500 TFYEDA-ALKKDIYRLAYSTNT-----VGYRFGGEYTGYYGSDDEKR 541  
Db 370 EYLSASALYRLVLSNLSFSLTSDASINRLNANLKDFAYPTRYSLTAFAGKYND----- 424  
Qy 542 AFGENSPTYKHKHCQSGIYEPVLK---KYGKKRANHVS--ISADFGDYFMPFAS--- 593  
Db 425 -----WLTASAVLATVINEEVROGSAANRRKLSPYVSAS-----FKPFASEEP 469  
Qy 594 -----YSRTHRMPNIQEMYSQIGDSGVHTALKPERANTWQFGNTYKGLLKQDDTLGL 648  
Db 470 RIRLFYKIDFRLPSFNDLYYGQVQ---NTNLKPESTTQYNLGL-TYSRSINELIPVVS 524  
Qy 649 KLVGYRSRIDNYIINVGKWDNLGNTPSVSVSTGLAYTQHRNFKDKVHKHGELELNY 708  
Db 525 TADAYNKVKDKIIAIPTK-----NLFIW-SMVNLG-----KVDIKGDIAGNI 567  
Qy 709 D---YGRFFTNLS--YAYQKSTQPTNFSASESPNNASKEDQLKQGVGLSRVSLPRDYG 763  
Db 568 SLQWFKLRVNLNSNYIYQRAL-----DMTEPGGKYKQ-----QIATPRVSG 611  
Qy 764 RLEVQ--TRW--LGNKLTGCMARYFGKSRATAEERYIDGTNGGNTSNVRQLGKRSTIQ 819  
Db 612 SQAGIETPWVNLVSYSLFSGKRYMLGNLR---ENRLDSYDHSVSVSRDLRIRNVNT 667  
Qy 820 TETLARQPLIFDFVAAEYEPKKNLIFRAEVKNLFDPRY 856  
Db 668 SLTV-----EVLNLDKNY 681

RESULT 14

RESULT 15  
US-09-584-501A-2  
; Sequence 2, Application US/09584501A

US-09-540-209B-9746  
; Sequence 9746, Application US/09540209B  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES  
; FILE REFERENCE: 2709.1001-001  
; CURRENT APPLICATION NUMBER: US/09/540,209B  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 10444  
; SEQ ID NO 9746  
; LENGTH: 783  
; TYPE: PRT  
; ORGANISM: B.fragilis  
US-09-540-209B-9746

Query Match 3.5%; Score 172; DB 5; Length 783;  
Best Local Similarity 20.7%; Pred. No. 4.9e-05;  
Matches 137; Conservative 84; Mismatches 258; Indels 184; Gaps 29;

Qy 34 EAQIQVLEDVHVAKRVPKDKKVFTRQDIDARAVSTRQDIFKSSSENLDNIVRSIPCAFTQQDKS 93  
Db 107 EPETQSLGEVVVTAKSEARQLREQAMPMSVLSMQLOQCTVSNVODVLSKTVGTVIRNTGG 166  
Qy 94 SGIVS-LNIRGDSGFGRVNTMVDGITOTFYSTSTDAGRAGSSQFGASVDSNFTAGLDVV 152  
Db 167 VGSRRSVVRGLEG-KRIGFFIDGSPMNDNSDFDIN-----DIPVDMIDRIEY 215  
Qy 153 KGSFSGSAGINSLAGSANL-----RTLGVDDVVQGNNTYGLLLKGLTGNTKGNAMA 205  
Db 216 KGVVPARFGSGVGAGVNVIREYPPKYLDASYTESFNTHKLSL-----VTRNI-- 266  
Qy 206 AIGARKWLESASGVLYGHSRRSVAQNYRVGGGQHIGNFGAEYLERRKORYVQ---E 262  
Db 267 ---ATKLEFGG--GGFVYS---DNNYKME-----SPEEGLIIKRNHDKKKLAVA 311  
Qy 263 GGLFENNSGKW-----ERDFORPYWKTKWYQKYND---POELQ---YI 301  
Db 312 GSLK---ARKWFDLABFEPVFIHTFKIOGIEYNIEKAHTYSDAFTFANKLEKENFLT 367  
Qy 302 EGHDKSWRENLAPOYDI--TPIDPSSLKQOQSAGNLFKL--EYDGVFNKYTAQFRLNLTKIG 358  
Db 368 EGLD--MESNLAYAYTVFHVWDTAARYNWDGTTYPAVSEYGEIGKWSARN----- 419  
Qy 359 SRKIINRYQFNGLSNLSYANLNLTAAYNSGRQKYPKGSKEFGW-----GL 405  
Db 420 EKHTITHKHLNYYVNNHNSINLSLFSFASGHPKDDLKNKVVGKYNFRSTMASWIAGL 479  
Qy 406 LKDEETYNNAKILDN-----NT-----ATFRLPRETELOTLGTFNYFNE 446  
Db 480 GYDFRTDNDIFLNSLNKYMYGNMTHMSSIMSSEAEKVDMLKRDGFSNALRYRFTPDF 539  
Qy 447 YGK-----NRFPEELGLFDPDQDNGLYSLGRFGDKGLLPQKSTIVQAGSQYFNT 500  
Db 540 MGKLSVGDVRLPAESELLGDG-----YTAPSGN--LLPERNVS-----NL 580  
Qy 501 FYFDAALKKDIYRLNYSNTVGYRFGGEYTGYYGSDDEKFRAGFENSPTYKHKHCQSGI 560  
Db 581 GFLDLRTEKDNALQVEVNT-----FYGVLENMIRFTG-----GY 615  
Qy 561 YEPVLKYYGKKRANHNSVSIADFGDYFMPFASYS-----RTH-----RM 600  
Db 616 LQSOYQNFGRMKRTLGVEVEVKADLTHWLYGYCNMTYQDLRVRKFEPNTHITNPTKGRM 675  
Qy 601 PNI 603  
Db 676 PNI 678

```

; GENERAL INFORMATION:
; APPLICANT: Lo, Reggie Y.C.
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Potter, Andrew A.
; TITLE OF INVENTION: TRANSFERRIN BINDING PROTEINS OF
; TITLE OF INVENTION: PASTEURELLA HAEMOLYTICA AND VACCINES CONTAINING THE SAME
; FILE REFERENCE: A34763 021645.0106
; CURRENT APPLICATION NUMBER: US/09/584,501A
; CURRENT FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: 08/753,759
; PRIOR FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: CA 2,164,274
; PRIOR FILING DATE: 1995-12-01
; PRIOR APPLICATION NUMBER: 60/008,569
; PRIOR FILING DATE: 1995-12-01
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Pasteurella haemolytica
US-09-584-501A-2

```

Query Match	3.4%	Score 166.5	DB 5	Length 930
Best local similarity	19.5%	Pred. No. 0.00017		
Matches 220	Conservative 139	Mismatches 354	Indels 415	Gaps 62
QY	5	FRUKPICFYLMGVMLYHHSYAEDAGRAGSEAOIQVLEDEVHVHAKRVPKDKKVFDDARAVS	64	
DB	8	FRYSPVALTVLFAL--SHSY-----GAATENKKEENNDLAV-----LDEVIVTESHYAH	55	
QY	65	TROD-----IPKSSENLDNIYRS-IPGA--FTQODKSSGIVSLNIRGDSGF-----GRV	110	
DB	56	ERONEVTGLGVKVKVKNYHEMSKNQILGRLDTRYDPGISVVEQGRGASSGYAIRGVDKNRV	115	
QY	111	NTWVDGITQ--TFYSTSDAGRAGGSGQFASVDNSNFIAGLDVVVKGFSFGSAGINSIAGS	168	
DB	116	SLLDVGLPQASHSYHTLGSDAN--GGAIN--ETIYENIRSTELSKASSAEYSGAHHGA	170	
QY	169	ANLRTGLGVDDVQGNNTYIGLLKGLTGTNSTKGMAAAIGARKWLESGASGVGLYGHRR	228	
DB	171	IGFRTKDAQDIIKEGOWH-----LDSKTSYASKNSHFLQ	205	
QY	229	SVAQNYRVGGGGHIGNFGAEXYLERRKQRYFVQEGGLKFNSNGSKWBRDFORPYWKTWY	288	
DB	206	SIAAAGEAG-----GFEALVIATHRGKE--TKIHSEANKLHNIR--ITGFE	250	
QY	289	QKYNDPQ-----ELQYIE-----GHDKSWRENLA-----POYDITPIDPSS	325	
DB	251	NRVDFQIPHRLLEDLLIVEDCPTLDCTPRARVKLRNDFPVRTPEY--TPEERKQ	308	
QY	326	LKQ-----QSAGNLFKLEYDGVFNKYTAQF-----RDLNTKIG	358	
DB	309	LEQIPVTEOLSAQYTGKDRIAPNPLDYKNSVFMKFGYHFNSSHVYLGAILEDTKTRYD	368	
QY	359	SRKIINRNY--QFNYGLKSLNSYANLNLTAAYNSG-----RQYKPKSGKETGWGLKD	408	
DB	369	IRDQWPAYYTKDINLSLRNY-----VYEGDNILDLGLVFXPRIPYGLR-----	412	
QY	409	FETYNNAKILLDLNNTATFRPRETELQTTLGFNYFHNEYKGNRFPPEELGLFFDGPDDNG	468	
DB	413	---YSHVKFPD-----ERHHKRRLGFTYKYKP-ENNRWLDISKLSAD--KODIE	455	
QY	469	LYSYLGR-----FKDGKGLLPKQSTIVQ-----PAGSQYF	498	
DB	456	LYSLRHLHCSDYPVVDKNCRPTLDKSWSMYRTERNNYQEKHRYIHLEFFKALNAGOGVF	515	
QY	499	N-TFYDDAALKKDIYR--LNYSTNTVGYRFGGYTGYGSDDEFKRAFGNSPTYKXH--	553	
DB	516	NOTHKUNLGLGTFORFNSLMDHGMDTAQYTKGG-YTSTRG-----RGRLONPIYTRDPR	568	
QY	554	-----CNOSCGIY---EPLVKKYGKKRANNHNSVTSADFGDYFMFPFASYSRTHRPN	603	

Search completed: July 24, 2002, 09:01:09  
Job time: 425 sec

Db	569	SIETYLCNTRGDILNCEP-----RKIGDSHFVS-----F	600
Qy	604	QEMYFSGIDSGVHTALKPERANTWQFGFTYKGLLKODDTLGLKLVGYRSRIDYIHN	663
Db	601	RDVISEYVDVLGVRFDOHR-----FKSDDPWTLST-YRN-----	636
Qy	664	VYGKMWDLNGI-----PSWYSSTGLAYTIOHRNFKDKVHKHGFELNYD-YGRFETNLS	718
Db	637	-----WSWNGSIGITLKPTFYS--LSYRI-----SNGPRVPAFYELGYK-----	672
Qy	719	YAYOKSTOPTNFSDAESPNNASKEDOLKOGYGLSRVSALPRDYGRLEVTGTRWLGKHLPL	778
Db	673	-----RDHIGLKD-NEYVQRAQRSHQLE-----PEKSTNHEIGVSPFKQGFYL	714
Qy	779	GGAMRYFG--KSRATAERYIDGT-----NGGNTSNVRQLGKRSTKQETELARQPLIFD	831
Db	715	D--VSYERNNYKNMIATACKRIIQKSHCFYNYINIQDVALINGNLVAKFDLHGILSMPLD	772
Qy	832	-FYA--AYEPKKNLIFRAEVKNLFDTRYIDP-LDAGNDA--ATQ--RYVSFPDPKDKDE	882
Db	773	GFYSSVAIN-----RVKVK---ERKLTDSRLDSVNDPILDIAQPARVLGFGYDHPEE	822
Qy	883	E--VFC-----NADK-----TLCNGK	896
Db	823	KWGIQITITTSKAKNADEVAGTRHHGTHRYVDLGGKLTGSWYTHDITGYINXKNYTLRGCI	882
Qy	897	YGGTSKSVLT-----NPAR-----GRFTLITMSYKF	922
Db	883	YNVTNRRKYSTWESVROSGVNADVODRGSNVTYREGCAPGNFSLAFEMKF	930

Search completed: July 24, 2002, 09:01:09  
Job time: 425 sec



**THIS PAGE BLANK (USPTO)**



# INTERFERENCE

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:53:48 ; Search time 25.29 Seconds  
(without alignments)  
889.520 Million cell updates/sec

Title: US-09-762-926-6  
Perfect score: 4894  
Sequence: 1 MRSSFLKPICFYLMGVTLV.....SVLTNFARGRTFLITMSYKF 921

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 231628 seqs, 24425594 residues  
Total number of hits satisfying chosen parameters: 231628

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : Issued\_Patents\_AA:\*  
1: /cgn2\_6/ptodata/2/iaa/5A\_COMB.pep:\*  
2: /cgn2\_6/ptodata/2/iaa/5B\_COMB.pep:\*  
3: /cgn2\_6/ptodata/2/iaa/6A\_COMB.pep:\*  
4: /cgn2\_6/ptodata/2/iaa/6B\_COMB.pep:\*  
5: /cgn2\_6/ptodata/2/iaa/PCTUS\_COMB.pep:\*  
6: /cgn2\_6/ptodata/2/iaa/backfiles1.pep:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

## SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	276.5	5.6	915	3	US-08-613-009A-18
2	273.5	5.6	915	1	US-08-487-890A-96
3	273.5	5.6	915	2	US-08-363-124A-2
4	273.5	5.6	915	2	US-08-478-435-96
5	273.5	5.6	915	2	US-08-337-483-96
6	273.5	5.6	915	2	US-08-478-373-96
7	273.5	5.6	915	3	US-08-474-671-96
8	273.5	5.6	915	3	US-08-483-577A-96
9	273.5	5.6	915	4	US-08-897-438-96
10	273.5	5.6	915	4	US-08-637-654-96
11	269	5.5	908	1	US-08-487-890A-94
12	269	5.5	908	2	US-08-478-435-94
13	269	5.5	908	2	US-08-337-483-94
14	269	5.5	908	2	US-08-478-373-94
15	269	5.5	908	3	US-08-474-671-94
16	269	5.5	908	3	US-08-483-577A-94
17	269	5.5	908	3	US-08-448-194-4
18	269	5.5	908	3	US-08-613-009A-16
19	269	5.5	908	4	US-08-897-438-94
20	269	5.5	908	4	US-08-867-921-4
21	269	5.5	908	4	US-08-637-654-94
22	267.5	5.5	791	3	US-08-537-361E-4
23	264.5	5.4	791	4	US-08-817-707-4
24	261.5	5.3	790	4	US-08-817-707-6
25	256.5	5.2	911	1	US-08-487-890A-95
26	256.5	5.2	911	2	US-08-478-435-95
27	256.5	5.2	911	2	US-08-337-483-95

28	256.5	5.2	911	2	US-08-478-373-95	Sequence 95, Appl
29	256.5	5.2	911	3	US-08-474-671-95	Sequence 95, Appl
30	256.5	5.2	911	3	US-08-483-577A-95	Sequence 95, Appl
31	256.5	5.2	911	3	US-08-448-194-6	Sequence 6, Appl
32	256.5	5.2	911	3	US-08-613-009A-17	Sequence 17, Appl
33	256.5	5.2	911	4	US-08-897-438-95	Sequence 95, Appl
34	256.5	5.2	911	4	US-08-867-921-6	Sequence 6, Appl
35	256.5	5.2	911	4	US-08-637-654-95	Sequence 95, Appl
36	253.5	5.2	790	4	US-08-817-707-8	Sequence 8, Appl
37	253.5	5.2	909	2	US-08-363-124A-4	Sequence 2, Appl
38	252.5	5.2	884	1	US-08-066-167-2	Sequence 2, Appl
39	250.5	5.1	790	3	US-08-537-361E-8	Sequence 8, Appl
40	249	5.1	941	4	US-09-074-658-75	Sequence 75, Appl
41	247	5.0	792	3	US-08-537-361E-6	Sequence 6, Appl
42	246	5.0	792	1	US-08-326-670A-2	Sequence 2, Appl
43	246	5.0	792	3	US-08-990-470A-2	Sequence 2, Appl
44	246	5.0	792	4	US-08-817-707-2	Sequence 2, Appl
45	239.5	4.9	792	3	US-08-537-361E-2	Sequence 2, Appl

## ALIGNMENTS

RESULT 1  
US-08-613-009A-18  
; Sequence 18, Application US/08613009A  
; Patent No. 6090576  
; GENERAL INFORMATION:  
; APPLICANT: Myers, Lisa E  
; APPLICANT: Schryvers, Anthony B  
; APPLICANT: Harkness, Robin E  
; APPLICANT: Loosmore, Sheena M.  
; APPLICANT: Du, Run-Pan  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Klein, Michel H  
; TITLE OF INVENTION: Transferrin Receptor Genes of Moraxella  
; NUMBER OF SEQUENCES: 31  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Slim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.30  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/613,009A  
; FILING DATE: 08-MAR-1996  
; CLASSIFICATION: 435  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24973  
; REFERENCE/DOCKET NUMBER: 1038-542  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 18:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 915 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-613-009A-18

Query Match 5.6%; Score 276.5; DB 3; Length 915;  
Best Local Similarity 21.2%; Pred. No. 4.2e-14;  
Matches 221; Conservative 128; Mismatches 328; Indels 367; Gaps 57;

```
.QY 5 FRKPICFYLMGVTLHYHYSAEDAGRAGSEAOIQVLEDVHVAK--RVPKDKKVFDTDARA 62
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
7 FRNLICLSLMTAL---PAYAENV-QAG-QAOKQLDTIOVKAKKQKTRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQDK--SSGIIVSLNIRGDSGFGFVNTVMDGIT 118
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 VKTADTSLKEQVLDRLDTRYPGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA 117
QY 119 Q-TFYSTSTAG--RAGGSQFQASVDSNFIAGLVDVVKFSGAGINSLAGSANLRTLGL 175
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
118 QIOSYTAQAALGGFRTAGSGAINIEYENVKAVEISKGSNSVEQSGSALAGSVAFQTKT 177
QY 176 VDDVVOGNNYVGLLKL-GLTGTNSTKGNMAAIGARKWLESGASGVGLYGHRSRVAQNY 234
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
178 ADDVIGEGROWGTQSTAYSGKNRGLTQSLALAG----- 211
QY 235 RVGGG-----GQHIGNFAEYLERKQRYFVQEGALKFN-----SDSGKWERDLOROQ 282
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
212 RIGGAELLITGHAGEIRAHEAGR-----GYQSFNRLAPVDDGSKYAYFIVEE 263
QY 283 WKYPKYNKNQELQ-----KYIEGDK-----SWRENLAPOYDITPIDPSSLKQOS-----A 330
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
264 -----CRNGHEKCANPPKDVVEDKQRTVSTRDYTPNRFLA--DPLSYESRSLFRP 316
QY 331 GNLF--KLEVDGVENKYTAQ--FRDLNTKIGSKII-----NRYQFN--Y 370
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
317 GFRENKRHYIGGLERTQOTFTRDVIYPAFLTKAVFDANQKQASLRNGKYGAGNHKY 376
QY 371 GLSLNPTNLNLTAAYNSGRQKPKGSKFTGWGLLKDFETYNNAKILDLNNTATFRLPRE 430
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
377 G-----GLFTSGENNAFVGAAY-GTGVFYD-----E 401
QY 431 TELQTLGFFNYFHEYKCNRPPELGLFFD--GPDQDNL----- 468
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
402 THTSRYGLEVYNADKDTWADYRUSYDRQIGLGNHFOQTHCSADGSDKYCRPSADK 461
QY 469 -YSYLRFGKDGKLLPKQSTIVOPAGSOYFNTFYFDAALKKDIYLNYSNTVGY-RFGG 526
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
462 PFSY---YKSDRVYIGESHLLQAAPKSPDTAKI-----RHLNLYN-LGYDRFGS 508
QY 527 -----EYTGYYGSDDEKFRAGFENSP----- 547
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
509 NLRHQDY--YQSAN---RAYSKTTPQNNKKTSPNGREKNPWWYSIGRGNVVVTRQICL 563
QY 548 ---TYKKHCNRSR---GIPEVPLK-----YCKKRANHSVS----- 578
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
564 FGNNTYDCTPRSLNGSKSYAAVRDNLGRWADVAGLRYDYRSTHSDGYSVSTGTHRT 623
QY 579 ISADFGDYFMPFASYSRTH-----RMPNQEMYFSQIGDSGVHTALKPERANTWQFG-- 630
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
624 LSNWAGIVLKPADWLDITYRTSTGFRLPSPAEMYGWRSGDKIRAKVIDPEKSPNKEAGIV 683
QY 631 -----FNTYKKGILLKQDDTLGLKIVGYSRIDN-----YIH----- 661
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
684 FKGDGFLNLEASWFNAYRDLIVR-----GYEAQIKDGKQKQVKNPAYLNAQSARIT 734
QY 662 --NYVKGK--W--WDLNGDIP-SWVSSTGLAYT-IQRNFKDKVKHKGFELELNLDYGRF 712
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
735 GINILGKIDWGVWD---KLPEGWYST--FAYNRRVVRDIIKRRADRTDIQSHL-FD--- 784
QY 713 FTNLSYAYQSTOPTNFSDASESPNNAKEDQLKQGVGLSRVLSALPRDYGRLEVGTWILG 772
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
785 -----AIQPSRVV-----VSGSYDQPEGWGVNGMLT-----YSKAKEITELLG 823
QY 773 NKLTGLGAMRYFGKSIRATAERYIDTNGNTSNFPLQKRSTIKQETTLARQPLIFDFY 832
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
824 SRALLNG-----NSRNTKATSRRTRPWYIVDVS 851
QY 833 AAYEPKKNLIFRAEVKNLFDERRYI 856
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
852 GYTTVKKHFTLRAGVYNLLNHRVY 875
```

## RESULT 2

```
US-08-487-890A-96
: Sequence 96, Application US/08487890A
: Patent No. 5708149
: GENERAL INFORMATION:
: APPLICANT: Loomore, Sheena
: APPLICANT: Harliness, Robin
: APPLICANT: Schryvers, Anthony
: APPLICANT: Chong, Pele
: APPLICANT: Gray-Owen, Scott
: APPLICANT: Yang, Yan-ping
: APPLICANT: Mordin, Andrew
: APPLICANT: Klein, Michel
: TITLE OF INVENTION: Transferrin Receptor Genes
: NUMBER OF SEQUENCES: 147
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Sim & McBurney
: STREET: 6th Floor, 330 University Avenue
: CITY: Toronto
: STATE: Ontario
: COUNTRY: Canada
: ZIP: M5G 1R7
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.25
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/487,890A
: FILING DATE: 07-JUN-1993
: CLASSIFICATION: 435
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/175,116
: FILING DATE: 29-DEC-1993
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: US 08/148,968
: FILING DATE: 08-NOV-1993
: ATTORNEY/AGENT INFORMATION:
: NAME: Stewart, Michael I
: REGISTRATION NUMBER: 24,973
: REFERENCE/DOCKET NUMBER: 1038-466 MIS:1b
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (416) 595-1155
: TELEFAX: (416) 595-1163
: INFORMATION FOR SEQ ID NO: 96:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 915 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-487-890A-96
```

```
Query Match 5.6%; Score 273.5; DB 1; Length 915;
Best Local Similarity 21.3%; Pred. No. 7.5e-14;
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

QY 5 FRKPICFYLMGVTLHYHYSAEDAGRAGSEAOIQVLEDVHVAK--RVPKDKKVFDTDARA 62
Db 7 FRNLICLSLMTAL---PAYAENV-QAG-QAOKQLDTIOVKAKKQKTRDNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQDK--SSGIIVSLNIRGDSGFGFVNTVMDGIT 118
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
62 VKTADTSLKEQVLDRLDTRYPGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA 117
QY 119 Q-TFYSTSTAG--RAGGSQFQASVDSNFIAGLVDVVKFSGAGINSLAGSANLRTLGL 175
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
118 QIOSYTAQAALGGFRTAGSGAINIEYENVKAVEISKGSNSVEQSGSALAGSVAFQTKT 177
QY 176 VDDVVOGNNYVGLLKL-GLTGTNSTKGNMAAIGARKWLESGASGVGLYGHRSRVAQNY 235
Db 111 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
178 ADDVIGEGROWGT-----QSKTAYSGKNRGL-----TQSIALAGR 212
```

```
QY 236 VGGG-----GQIGNFGAEYLERRKORYFVOEGALKEN-----SDSGKWERDLQROOW 283
Db 213 IGAEALLIIRTRHAGEIRAEHAGR-----GVOSFNRLAPVDGSGKYAIFIVEEE- 263
QY 284 KYKPKYNNNOEL-----QKYIEGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AG 331
Db 264 ----CKNGGHECKKANPKDVGEDKQTVSTRDTGPNRFLA--DPLSYESRSWLFRPG 317
QY 332 NLF--KLEYDGVFNKTAQ---FRDLNFKIGSRKII-----NRRYQFN--YG 371
Db 318 FRENKRHYIGGILERTQOTFTDRMTVPAPLTKAVFDANQKAGSLRGNGKYAGNHKG 377
QY 372 LSLNPTNLNLTAAYNSGRQKYPKSGKGTGGLLKDFETYNNAKILDLNNTATFRPRET 431
Db 378 -----GLFTSGENNAVPVGAIEY-GTGVEFD-----ET 402
QY 432 ELQTTIGFNFHNEYGKNRPEELGLFFD--GPDQDNGL----- 468
Db 403 HTKSRYLEVYVYNADKDTWADYARLSYDRQGIGLGNHFOQTHCSADGSDKYCRPSADKP 462
QY 469 YSLGRFKGDKGLLPKSTIVQAGSQYFNTFFDAALKKDIYRLNYSNTVGY-RFGG- 526
Db 463 FSY---YKSDRVYIGESHKLLQAAFKKSFDTAKI-----RHNL SVN-LGYDRFGSN 509
QY 527 ---EXTGYGSDDEKRAFGENSP----- 547
Db 510 LRQDY--YQSAN---RAYSLKTPPNNGKKTSPNGREKNPYWVSGIRGNVVTQICLF 564
QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YGKKRANNHSVS-----I 579
Db 565 GNNTYDCTPRSNGKSYAAVRDYNRLGRWADVAGLRYDYRSTHSDGSGVSTGHTL 624
QY 580 SADFQGYFMPFASYSRTH-----RMPNIQMYFSQIGDSGVHTALKPERANTWQFG--- 630
Db 625 SWNAGIVLKPADWLDLTYRTSTGRFLPSFAEMYGWSRQKIKAVKIDPEKSFENKEAGIVF 684
QY 631 -----FNTYKGLLKQDDTLGLKLVGYRSRDN-----YIH----- 661
Db 685 KGFNGLEASFNNAVRLDLVR-----GYEAQIKQKGEQVGNPNAQASARITG 735
QY 662 -NVYK--W---WDLNGDIP-SWVSSTGLAYT-IQRNFKDKVHKHGFEELELNDYGRFF 713
Db 736 INILGKIDWGVND---KLEPGWYST--FAYNRVVRDINKRADRTDIQSHL-FD----- 784
QY 714 TNLAYAKSTQTPNFSADASESPNNASKEDQLKOGYLSRVSALPRDYGRLEVGTWRLGN 773
Db 785 -----AIQPSRYV-----VSGYDQPEGKGVNGMLT-----YSKAKETELLGS 824
QY 774 KLTIGGAMRYFGKSIRATAERYIDGTNGGNTSNFROLGKRSIKQETTLARQPLIFDYA 833
Db 825 RALLG-----NSRNTKATARRTPWYIVDVSG 852
QY 834 AYEPKKNLIFRAEVKNLFDRIYI 856
Db 853 YTVVKKHFTLRAGVYLLNHRIV 875
```

## RESULT 3

```
US-08-363-124A-2
; Sequence 2, Application US/08363124A
; Patent No. 5912336
; GENERAL INFORMATION:
; APPLICANT: Sparling, P. Frederick
; APPLICANT: Cornelissen, Cynthia N.
; TITLE OF INVENTION: Transferrin-Binding proteins from
; TITLE OF INVENTION: Neisseria Gonorrhea and Neisseria Meningitidis
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: ImClone Systems Incorporated
; STREET: 180 Varick Street
; CITY: New York
; STATE: New York
; COUNTRY: USA
```

```
; ZIP: 10014
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/363,124A
; FILING DATE: 23-DEC-1994
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/124,254
; FILING DATE: 20-SEP-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/973,336
; FILING DATE: 05-NOV-1992
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/572,187
; FILING DATE: 23-AUG-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Gallagher, Thomas C.
; REGISTRATION NUMBER: 37,066
; REFERENCE/DOCKET NUMBER: SPA-1-PDC
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 645-1405
; TELEFAX: (212) 645-2054
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 915 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-363-124A-2
```

Query Match 5.6%; Score 273.5; DB 2; Length 915;

Best Local Similarity 21.3%; Pred. No. 7.5e-14; Indels 365; Gaps 57;

Matches 222; Conservative 129; Mismatches 327;

```
QY 5 FRLLPICFLYLMGVTLYHYSYAEDAGRAGEAQIQLVEDVHVAK--RVPKDKVFTDARA 62
Db 7 FRNLILSLMTAL---PAYAENV-QAG-QAQEKQLDTIQVAKKOKTRDRNEVTGLGKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQDK--SSGIVSLNIRGDSGFRVNTWVDGIT 118
Db 62 VKTADTLKSKQVLDIRLDYDFGIAVVEQGRASSG---YSIRG-MDKNRVSLTVDGLA 117
QY 119 Q-TFYSPTDAG--RAGGSSQFGASVDSNFTAGLDVYKGSFGSGAGINSLAGSANLRTL 175
Db 118 QIOSYTAQAALGGTRTAGSSGAINIEYENVKAVEISKGSNSVEQSGGALAGSAFQTKT 177
QY 176 VDDVVOGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESGASGVLYLGHSSRVAQNYR 235
Db 178 ADVIGEGROWGI-----QSKTAYSKNRL-----TQSIALAGR 212
QY 236 VGGG-----GQIGNFGAEYLERRKORYFVOEGALKFN-----SDSGKWERDLQROOW 283
Db 213 IGAEALLIIRTRHAGEIRAEHAGR-----GVOSFNRLAPVDGSGKYAIFIVEEE- 263
QY 284 KYKPKYNNNOEL-----QKYIEGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AG 331
Db 264 ----CKNGGHECKKANPKDVGEDKQTVSTRDTGPNRFLA--DPLSYESRSWLFRPG 317
QY 332 NLF--KLEYDGVFNKTAQ---FRDLNFKIGSRKII-----NRRYQFN--YG 371
Db 318 FRENKRHYIGGILERTQOTFTDRMTVPAPLTKAVFDANQKAGSLRGNGKYAGNHKG 377
QY 372 LSLNPTNLNLTAAYNSGRQKYPKSGKGTGGLLKDFETYNNAKILDLNNTATFRPRET 431
Db 378 -----GLFTSGENNAVPVGAIEY-GTGVEFD-----ET 402
QY 432 ELQTTIGFNFHNEYGKNRPEELGLFFD--GPDQDNGL----- 468
```

Db 403 HTKSRYLEVYVTTNADKTDWADYARLSYDRQIGLDNHFQOHCSDGSKYCRPSADKP 462  
QY 469 YSLGRFKGDKLLPQKSTVQVAGSYFNTFYFDAALKKDIYRLNVTNTVGY-RFGG- 526  
Db 463 FSY---YKSDRVYIGESHKLLQAAFKSFDTAKI-----RHNLSVN-LGYDRFGSN 509  
QY 527 ----EYTGYSDDDEFKRAFGENSP----- 547  
Db 510 LRHODY--YYQSAN---RAYSLKTPPNQNGKKTSPNGREKNPYVWSTGRGNVTVTRQICLF 564  
QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YKKRANNHSVS-----I 579  
Db 565 GNNYTDCTPRSLNGKSYAAVRDNVRLGRWADVAGLRYDYRSTHSDGCVSTGTHRTL 624  
QY 580 SADGDFYMFASFYSRTH-----RMPNIQMFYSQIGDSGVHTALKPERANTWQFQ--- 630  
Db 625 SWNAGIVLKPADWLDLTYRTSTGRFLPSFAEMYGWSGDKIKAVKIDPEKSFNKEAGIVF 684  
QY 631 -----FNTYKGLKLQDDTLGLKLVGYSRIDN-----YIH----- 661  
Db 685 KGDFGNLEASWFNAYRDLIVR-----GYEAQIKDGKEQVKNPYNQAQSAKITG 735  
QY 662 -NYGK--W---WDLNGDIP-SWVSSGTLAYT-IOHENFKDKVHKHGFELNAYDYGRFF 713  
Db 736 INILGKIDWNGWD---KLPEGWYST--FAYNRVRVDIKKRADRTDIQSHL-FD----- 784  
QY 714 TNLAYQKSTQPTNFSDASESPNNAKEDQLKOGYGLSRVSALPRDYGRLEVGTRWLG 773  
Db 785 -----AIQPSRYV-----VGSYDQPEKGWGVNGLT-----YSRAKETEILGJS 824  
QY 774 KLTGGAMRYFGKSIRATAERYIDGTNGGTSNFRQLGRKRSIKQETTLARQPLIFDYA 833  
Db 825 RALLNG-----NSRNTKATARRTRPWYIVDVSG 852  
QY 834 AYEPPKNLIFRAEVKNLFDRRYI 856  
Db 853 YYTVKKHFTLRAGVYNLLNHRIV 875

RESULT 4  
US-08-478-435-96  
; Sequence 96, Application US/08478435  
; Patent No. 5922323  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena  
; APPLICANT: Harkness, Robin  
; APPLICANT: Schryvers, Anthony  
; APPLICANT: Chong, Pele  
; APPLICANT: Gray-Owen, Scott  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Murdin, Andrew  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Transferrin Receptor Genes  
; NUMBER OF SEQUENCES: 147  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,435  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/337,483  
; FILING DATE: 08-NOV-1994

; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/175,116  
; FILING DATE: 29-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/148,968  
; FILING DATE: 08-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 96:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 915 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-478-435-96

Query Match 5.6%; Score 273.5; DB 2; Length 915;  
Best Local Similarity 21.3%; Pred. No. 7.5e-14;  
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

QY 5 FRKPCIFGLMGTLYHYSAEDAGRAGSEAOIQVLEDVHVAK--RVPKDKKVFETDARA 62  
Db 7 FRNIICLSLMTAL---PAYAENV-QAG-OAQEKQLDTIOVKAKKQKTRDNEVTGLGKL 61  
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGVRVMTVDGIT 118  
Db 62 VKTADTLSKEQVLDIRLDTRYDPCIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA 117  
QY 119 Q-TFYSTSTTAG--RAGSSQFCASVDNFIAGLDVVVKGSFSGSAGINSLAGSANLRTLG 175  
Db 118 QIOSYTAQAALGCTRTAGSSGAINIEYENVKAVEISKSGNSVEQSGGALAGSVAFOTKT 177  
QY 176 VDDVQGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASVGLVYHSRVSVAQNYR 235  
Db 178 ADDVIGEGROWGI-----QSKTAYSGKNRGL-----TQSIALAGR 212  
QY 236 VGGG-----GQHIGNFGAEYLERRKQRYVQEGALKFN-----SDSGKWERDLQKQW 283  
Db 213 IGGAEALLIRTHAGEIRAEHAAGR-----GVQSFNRLAPVDDGSKYAFVIVEE- 263  
QY 284 KYPKYKNYNNQEL---QKYLEGHDK---SWRENLAPOVDITPIDPSSLKQOS-----AG 331  
Db 264 ----CKNGGHEKCKANPKKDVGDGDKRQTVSTRDYTGPNRFLA--DPLSYESKSWLFRPG 317  
QY 332 NLF--KLEYDGVFNKYTAQ---FRDLNLTIGSRKII-----NRRYQFN--YG 371  
Db 318 FRENRKHYIGGILERTQQTFTDTRDMTVPAFLTKAVFDANQKQAGSLRGNGKYAGNHKYG 377  
QY 372 LSLNPTYNLNLTAAYNSGRQKYPKSKFTGWLGLLKDFETYNNAKILDLNNTATFRLPRET 431  
Db 378 -----GLFTSGENNAIPVGAEY-GTGVFYD-----ET 402  
QY 432 ELQTTLGFNFHNEYKKNRPEELGLFFD--GPDQDNGL----- 458  
Db 403 HTKSRYLEVYVTTNADKTDWADYARLSYDRQIGLDNHFQOHCSDGSKYCRPSADKP 462  
QY 469 YSLGRFKGDKLLPQKSTVQVAGSYFNTFYFDAALKKDIYRLNVTNTVGY-RFGG- 526  
Db 463 FSY---YKSDRVYIGESHKLLQAAFKSFDTAKI-----RHNLSVN-LGYDRFGSN 509  
QY 527 ----EYTGYSDDDEFKRAFGENSP----- 547  
Db 510 LRHODY--YYQSAN---RAYSLKTPPNQNGKKTSPNGREKNPYVWSTGRGNVTVTRQICLF 564  
QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YKKRANNHSVS-----I 579









STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/483,577A  
FILING DATE: 07-JUN-1995  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/337,483  
FILING DATE: 08-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-511  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 915 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-483-577A-96

Query Match 5.6%; Score 273.5; DB 3; Length 915;  
Best Local Similarity 21.3%; Pred. No. 7.5e-14;  
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

QY 5 FLRLPICFLMGVLYHYSAEDAGRAGSEAQIOVLEVDHVHAK--RVPKDKVFTDARA 62  
DB 7 FLRLNLCSLMTAL---PAYAENY-QAG-QAQRKQLDTIOVKAKKQKTRRDNEVTGLGKL 61  
QY 63 VSTRDIFKSSSE-NLDNLVRSIPG-AFTODK--SSGIVSLNIRGDSGFGRVNTMVDGIT 118  
DB 62 VKTADTISKEOVLDIRLTRYDPGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA 117  
QY 119 Q-TFYSTSTDAG--RAGSSQFGASVDSNFTAGLDVVYKGFSGSAGINSLAGSANLRTL 175  
DB 118 QIQSYTAQAALGGTRTAGSGAINIEIEYENVKAVEISKGSNSVEQSGGALAGSVAFTKT 177  
QY 176 VDDVVQGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESGASVGVLYGHSRRSVAQNYR 235  
DB 178 ADDVIGEGROWGI-----QSKTAYSGKNRGL-----TQSIALAGR 212  
QY 236 VGGG-----GQHIGNFGAEYLERRKQRFVQEGALKFN-----SDSGKWERDLQROQW 283  
DB 213 IGGAEALLIRTRHAGEIRAHEAAGR-----GVQSFNRLAPVDGSKYAYFIVEEE- 263  
QY 284 KYKPYKNYNQOEL-----OKYTEGDK-----SWRENLAPOYDITPIDPSSLKQOS-----AG 331  
DB 264 ---CKNGGHEKCKANPKDVGEDKQTVSTRDYTGPNRFLA--DPLSVESRSWLFPRG 317  
QY 332 NLF--KLEYDGVFNKYTAQ---FDLNTKIGSRKII-----NRNQFN--YG 371  
DB 318 FRFENKRRHYIGGILERTQQTFTDMDTVPAFLTKAVFDANOKQAGSLRGKYGAGNKHG 377  
QY 372 LSLNPNYLNLTAAYNAGROKYPKGSKFTGWGLLKDFETYNNAKILDLNNTATFRLPRET 431  
DB 378 -----GLFTSGENNAVPYGAEEY-GTGVFYD-----ET 402

QY 432 ELQTTLGFNFYHNEYGKNRFPPEELGLFFD--GPDDQDNL----- 468  
DB 403 HTKSRYGLEIVYTNADKDTWADYARLSYDRQIGLDNHFQOOTHCSADGSKYCRPSADKP 462  
QY 469 YSILGRFKGDKGLLPQKSTIVQAGSQYFNTFFYDAALKKDIYRLNYSNTNFGY--RFGG- 526  
DB 463 FSY---YKSDRVYIGESHKLLQAFAFKKSFTAKI-----RHNL SVN--LGYDRFGSN 509  
QY 527 ----EYTGYYGSDDEFKRAFGENSP----- 547  
DB 510 LRHODY--YYOSAN---RAYSLKTPPONNGKKTSPNGREKNPYWVSGRGNVTVTRQICLF 564  
QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YGKRRANNHSVS-----I 579  
DB 565 GNNYTTDCTPRSINGKSYAAVRDNVRLGRWADYAGLRYDYRSTHSDGGSVSTGTHRTL 624  
QY 580 SADFGDYFMPFASYSRTH-----RMPNIOEMYSQIGDSGVHTALPERANTWQFG--- 630  
DB 625 SWNAGIVLKPADWLDLTFTSTGFRLPFAEMYGWRSGDKIKAVKIDPEKSFNKEAGIVF 684  
QY 631 -----FNTYKKGILLKQDDTLGLKLVGYBSRIDN-----YIH----- 661  
DB 685 KGDFGNLEASWFNNAYRDLIVR-----GYEAQIKDKGEQVYGNPAYLNAOSARITG 735  
QY 662 -NHYGK--W---WDLNGDIP--SWVSSSTGLAYT-IQHNFKDKVHKHGFLELNDYDGRFF 713  
DB 736 INILGKIDWNGVMD--KLPEGWYST--FAYNRVVRDIKKRADRTDIQSHL-FD----- 784  
QY 714 TNLISAYOKSTQPTNPSDASESPNNASKEDOLKOGYGLSRVSALPRDYGRLEVTRWLN 773  
DB 785 -----AIQPSRYV-----VSGSYDQPEGKWGVNGMLT-----YSKAKEITELLS 824  
QY 774 KLTILGAMRYFGKSIRATAEERYIDGTNGNTSNFRQLGKRSIKQETLARQPLIFDYA 833  
DB 825 RALLNG-----NSRNTKATARRTPWTIVDVSG 852  
QY 834 AYEPPKKNLIFRAEYKKNLFDRIYI 856  
DB 853 YITVKKHFTLRAGVYNLLNHRV 875  
RESULT 9  
US-08-897-438-96  
; Sequence 96, Application US/08897438  
; Patent No. 6262016  
; GENERAL INFORMATION:  
; APPLICANT: Loomore, Sheena  
; APPLICANT: Harkness, Robin  
; APPLICANT: Schryvers, Anthony  
; APPLICANT: Chong, Pele  
; APPLICANT: Gray-Owen, Scott  
; APPLICANT: Yang, Yan-ping  
; APPLICANT: Mordin, Andrew  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Transferrin Receptor Genes  
; NUMBER OF SEQUENCES: 160  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: 6th Floor, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/897,438  
; FILING DATE:  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:

APPLICATION NUMBER: US 08/483,577  
FILING DATE: 07-JUN-1995  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/337,483  
FILING DATE: 08-NOV-1994  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-720  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 96:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 915 amino acids,  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-897-438-96

Query Match 5.6%; Score 273.5; DB 4; Length 915;  
Best Local Similarity 21.3%; Pred. No. 7.5e-14;  
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

QY 5 FRLLPICFYLGMVLYHYSAEADAGRAGSEAOIQVLEDVHVAK--RVPKDKKVFVDARA 62  
DB 7 FRLNLCISLMTAL---PAYAENV-QAG-QAQBKQDLTIQVAKKQKTRDNEVTGLGKL 61  
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGISLNIIRGDSGFGVRVNTWDGIT 118  
DB 62 VKTADTSLKEQVLDIRDLTRYDGIADVVEQGRGASSG---YSIRG-MDKNRVSLTVDGCLA 117  
QY 119 Q-TFYSTSDAG--RAGSGQFQGSVDNFIAGLDVVKVGFSGSAGINSIAGSANLRTLG 175  
DB 118 QIQSYTAQAALGGTGTAGSSGAINIEYENVKAVEISKGSNSVEQSGALAGSVAFTKT 177  
QY 176 VDDVQGNNTYTGILLKCLTGTNSTKGNMAAIGARKWLESGASGVLYGHSRVSVAQNYR 235  
DB 178 ADDVIGEGWGI-----QSKTAYSGKNRGL-----TQSIALAGR 212  
QY 236 VGGG-----GQHIGNFGAEYLERRKORYVQEGALKEN-----SDSGKWERDLQROOW 283  
DB 213 IGAEALLIRTRHAGEIRAEHAGR-----GVQSFNRLAPVDDGSKYAFIVEEE- 263  
QY 284 KYPKYNYNNOEL-----QKYIEGHDK---SWRENLAPOVDITPIDPSSLKQOS---AG 331  
DB 264 ---CKNGHEKCKANPKDVGDGKQRTVSTRDYTCNRFLEA--DPLSYESRSWLFPRG 317  
QY 332 NLP--KLEYDGVNKYTAQ---FRDLMTKIGSRKII-----NRNFOFN--YG 371  
DB 318 FREENKHYIGGILERTQOTFTDRMTVPALTKAVFDANQKQAGSLRGNGKYAGNHKY 377  
QY 372 LSLNPTNLNLTAAYNSGRQKPKGSKFTGWLKDPETYNNAKILLDNLNTATFRLPRET 431  
DB 378 -----GLFTSGENNAVPVGAAY-GTGVFYD-----ET 402  
QY 432 ELQTTLGFNFHNEYGNRPPEELGLFFD--GPDQDQNL----- 468  
DB 403 HTKSRVGLVYVYNADKTDWADYARLSYDRQIGIGLDNHFQOHTCSADGSKYCRPSADKP 462  
QY 469 YSYLGRPKGDKLLPKQSTIVPAGSOYFNTFFYDAAKKDIYRLNYSNTVGY-RFGG- 526  
DB 463 FSY---YKSDRVYIGESHKLLQAAFKKSFDTAKI-----RHNLSVN-LGYDRFGSN 509  
QY 527 ---EYTGYYGSDDEKPRAGENSP----- 547

DB 510 LRHQDY--YYQSAN---RAYSLKTPPONNGKKTSPNGREKNPYWVSIGRNVVTRQICLF 564  
QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YKRRANNHSVS-----I 579  
DB 565 GNNTYTDCTPRSINGKSYAAVRDNVRLGRWADVAGLGRDYDRSTHSDGDSVSTGTHRTL 624  
QY 580 SADFGDYFMPFASYSRTH-----RMPNIQEMYFSQICDSGVHTALAPERANTWQFG--- 630  
DB 625 SWNAGIVLKPADWLDTYRTSTGTRLPFSFAEMYGWSRGDKIKAVKIDPEKSFNKEAIVF 684  
QY 631 -----FNTYKGLLQDDTLGLVGYRSRDN-----YIH----- 661  
DB 685 KGDFGNLEASWFNNAYRDLIVR-----GYEAQIKDKGEQVKNPAYLNAQSARITG 735  
QY 662 -NVYKG--W---WDLNGDIP-SWVSTGLAYT-IQHRNFKDKVHKHGFELNLYDYGREF 713  
DB 736 INILGKIDWNGVWD---KLPEGWYST--FAYNRVRVRDIKKRADRTDIQSHL-FD----- 784  
QY 714 TNLASYAVKSTOPTNFSASESPNNASKEDOLKQGYGLSRVSALPRDYGRLEVGCTRWLGN 773  
DB 785 -----AIQPSRYV-----VGSYDQPEKGKVGWNGMLT-----YSKAKEITELLS 824  
QY 774 KLTILGAMRYFGKSIRATABERYIDGTNGGNTSNFROLGRKRSIKQETLARQPLIFDYA 833  
DB 825 RALLNG-----NSRNTKATARTRPWIVDVSG 852  
QY 834 AYEPPKKNLIFRAEVKNLFDRIYI 856  
DB 853 YTVVKKHFTLRAGVYNLLNHRIV 875  
RESULT 10  
US-08-637-654-96  
Sequence 96, Application US/08637654  
Patent No. 6358727  
GENERAL INFORMATION:  
APPLICANT: Loosmore, Sheena M  
APPLICANT: Harkness, Robin E  
APPLICANT: Schryvers, Anthony B  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-ping  
APPLICANT: Mardin, Andrew D  
APPLICANT: Klein, Michel H  
TITLE OF INVENTION: HAEMOPHILUS TRANSFERIN RECEPTOR GENES  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: PatentIn Release #1.0, Version #1.30  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/637,654  
FILING DATE: 05-AUG-1996  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
PRIOR APPLICATION NUMBER: PCT/CA94/00616  
FILING DATE:  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-595  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 96:

SEQUENCE CHARACTERISTICS:  
LENGTH: 915 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-637-654-96

Query Match 5.6%; Score 273.5; DB 4; Length 915;  
Best Local Similarity 21.3%; Pred. No. 7.5e-14;  
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

QY 5 FRLKPICFYLMGVLYHYSAEDAGRAGSEAQIQVLEDVHVAK--RVPKDKKVFDTDARA 62  
Db 7 FRLNILCLSLMTAL---PAYAENV-QAG-QAQEQQLDTIQVAKAKKQKTRRDNEVTGLGKL 61  
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQDK--SSGIIVSLNIRGDSGFRVNTWVDCIT 118  
Db 62 VKTADTLSEKQVLDIRLTRYDPIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA 117  
QY 119 Q-TFYSTSTDAQ--RAGSSOFGASVDSNFTAGLDVYVKGFSGAGINSLAGSANLRTL 175  
Db 118 QIQSYTAQAALGGTRTAGSSGAINIEYENVKAVEISKGSNSVEQSGALAGSVAFOKT 177  
QY 176 VDDVVQGNNTYGLLLKGLTGTNSTKGNAMAIGARKWLESASGVLYLGHSSRRSVAQNYR 235  
Db 178 ADDVIGEGROWGI-----QSKTAYSGKNRGL-----TQSIALAGR 212  
QY 236 VGGG-----GQHIGNFGAEYLERRQYFVQEGALKEN-----SDSGKWERDLORQOW 283  
Db 213 IGGAEALLIRTRHAGEIRAEAGR-----GVQSFNRLAPVDDGSIYAFIVEEE- 263  
QY 284 KYKPYKNNOEL-----QKYEIGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AG 331  
Db 264 ---CKNGHEKCKANPKDVGDKQTVSTRDYTGPNRFLA--DPLSYESRSWLFRPG 317  
QY 332 NLF--KLEYGVFNKYTAQ---FRDLNKTIGSRKII-----NRNYQFN--YG 371  
Db 318 FRFENKRYHIGGILERTQOTFTDMDTVPAFLTKAVFDANOKOAGSLRGKYGAGNHYG 377  
QY 372 LSLNPYTNLNTAAYNSGRQKYPKSGKFTGTGLLKFDETYNNAKILDNLNTATFRLPRET 431  
Db 378 -----GLFTSGENNAVPGAAY-GTGVPYD-----ET 402  
QY 432 ELQTTLGFNFHNYGKNRPFPEELGLFFD--GPDQDNL----- 468  
Db 403 HTKSRYGLEYYVTNADKTDWADYARLSYDRQIGLDNHFQOHCSDGSDKYCRPSADKP 462  
QY 469 YSYLGRKPKDKGLLPQKSTIVQAGSOYFNTFFYDAALKKDIYRLNYSNTVGY-RFGG- 526  
Db 463 FSY---YKSDRVYIGESHKLLQAAFKKSFTAKI-----RHNLSVN-LGYDREFSN 509  
QY 527 ---EYTGYYGSDDEKFRAGFENSP----- 547  
Db 510 LRHQDY--YQSAN---RAYSLKTPPONNGKKTSPNGREKNPYNWVSGRNVVTRQICLF 564  
QY 548 ---TYKKHNRSC---GIYEPVLKK-----YKGRANNHVS-----I 579  
Db 565 GNNTYDCTPRSINGKSYAAVRNVRNLGRWADVAGLRYDYRSTHSDGSGVSTGTHRTL 624  
QY 580 SADGCDYFMPFASVSRTH-----RMPNIOEMYPFQSDGSGVHTALKAPERANTWQFG--- 630  
Db 625 SWNAGIVLKPADWLIDLTRYSTGFRLPFAEYMGWRSGDKIKAVKIDPEKSFNKEAGIVF 684  
QY 631 -----FNYYKGLLKQDDTLGLKLGVYRSRDN-----YIH----- 661  
Db 685 KGDFNLEASWNNAYRDLIVR-----GYEAQIKGCKEQVKGNPYNLAQASARITG 735  
QY 662 -NVYK--W---WDLNGDIP--SWYSSTGLAYT-IOHRNFKDKYVKKHGFEELELVNDYGRFF 713  
Db 736 INILGKIDWNGVMD---KLEPGWYST--FAYNRVRVEDIKKRDRTDIQSHL-FD----- 784  
QY 714 TNLAYQKSTQPTNFSDESSEPNNAKEDQLKOGYGLSRVSALPRDYGRLEVGTWRLGN 773

Db 785 -----AIQPSRYV-----VSGGYDQPEGKWGVNGMLT-----YSKAKEITELLS 824  
QY 774 KLTLLGGAMRYFGKSIRATAEERYIDGTNGNTSNFRQLGKRSIKQOTETLARQPLIFDEYA 833  
Db 825 RALLNG-----NSRNTKATARRTRPWYIVDVSG 852  
QY 834 AYEPPKKNLIFRAEVKNLFDRIYI 856  
Db 853 YYTVKKHFTLRAGVYNLLNRYV 875

RESULT 11  
US-08-487-890A-94  
Sequence 94, Application US/08487890A  
Patent No. 5708149  
GENERAL INFORMATION:  
APPLICANT: Loomore, Sheena  
APPLICANT: Harkness, Robin  
APPLICANT: Schryvers, Anthony  
APPLICANT: Chong, Pele  
APPLICANT: Gray-Owen, Scott  
APPLICANT: Yang, Yan-ping  
APPLICANT: Murdin, Andrew  
APPLICANT: Klein, Michel  
TITLE OF INVENTION: Transferrin Receptor Genes  
NUMBER OF SEQUENCES: 147  
CORRESPONDENCE ADDRESS:  
ADDRESSEE: Sim & McBurney  
STREET: 6th Floor, 330 University Avenue  
CITY: Toronto  
STATE: Ontario  
COUNTRY: Canada  
ZIP: M5G 1R7  
COMPUTER READABLE FORM:  
MEDIUM TYPE: Floppy disk  
COMPUTER: IBM PC compatible  
OPERATING SYSTEM: PC-DOS/MS-DOS  
SOFTWARE: Patent In Release #1.0, Version #1.25  
CURRENT APPLICATION DATA:  
APPLICATION NUMBER: US/08/487,890A  
FILING DATE: 07-JUN-1993  
CLASSIFICATION: 435  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/175,116  
FILING DATE: 29-DEC-1993  
PRIOR APPLICATION DATA:  
APPLICATION NUMBER: US 08/148,968  
FILING DATE: 08-NOV-1993  
ATTORNEY/AGENT INFORMATION:  
NAME: Stewart, Michael I  
REGISTRATION NUMBER: 24,973  
REFERENCE/DOCKET NUMBER: 1038-466 MIS:jb  
TELECOMMUNICATION INFORMATION:  
TELEPHONE: (416) 595-1155  
TELEFAX: (416) 595-1163  
INFORMATION FOR SEQ ID NO: 94:  
SEQUENCE CHARACTERISTICS:  
LENGTH: 908 amino acids  
TYPE: amino acid  
STRANDEDNESS: single  
TOPOLOGY: linear  
US-08-487-890A-94

Query Match 5.5%; Score 269; DB 1; Length 908;  
Best Local Similarity 21.2%; Pred. No. 1.7e-13;  
Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;

QY 5 FRLKPICFYLMGVLYHYSAEDAGRAGSEAQIQVLEDVHVAK--RVPKDKKVFDTDARA 62  
Db 7 FRLNILCLSLMTALPV---YAENV--QAEQAKQQLDTIQVAKAKKQKTRRDNEVTGLGKL 61

```
QY 63 VSTRQDIFKSSE-NLDNIVRSIPG-AFTQDK--SSGIYSLNIRGDSGFRVNTWVDGIT 118
Db 62 VKSSDLSKQVNLNLDLTRYDPIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGVS 117
QY 119 Q-TFYSTSDAG--RAGGSSQFGASVDSNFIAGLDVVYKGSFGSAGINSLAGSANLRTLG 175
Db 118 QIOSYTAQAALGCTRTAGSGGAINIEYENKAVEISKGSNSSEYNGALAGSVAFOTKT 177
QY 176 VDDVVOGNTTYGLLK-GLTGNTSTKGNMAAIGARKWLESASGVLYGHRSRVAQNY 234
Db 178 AADIIEGKGWGIQSKTAYSGKDHALTQSLALAGR---SGGAELIITYTKRGRETHAH 233
QY 235 RVGGGQHIGNECAEYLERKQ---RYFVOEGAL-----KFNSD-SGKWER-- 276
Db 234 KDAGKG--VVSFNRVLDEDKGGGQYRYFIVEEHCNGYAAACKNKLKEDASVKDERKT 291
QY 277 -----DLQROQWKYKP-----YKNYNQELQKYIEGHDKSWRENLAPO 314
Db 292 VSTQDYTGSNRLANPLEYGSQSLFRPGWHLDNRHVGVAVLERTQOTFTD--RDMTVPA 349
QY 315 YDTPTD--PSSLK-----QOSAGNLFKLEY-DGVFNKYTAQFRDLNLT 355
Db 350 Y-FTSEDYVPSLKLGLKYSKSGDNKAERLFVQGGESTLQIGIGYGTGVF-----YDERHTK 402
QY 356 IGSRKIIINRYQFNGLSLNPTNLNLTAAYNRSQKYPKGSFTGWLLKDEFTYNNAK 415
Db 403 -----NR-----YGVGY-VYHNADKDTWADYARLSYDRQG----- 431
QY 416 ILDNNTATPRLPRETELQTLTGFNTFNHNEYGKNRPEELGLFFDGDODNGLYSYLGRF 475
Db 432 -IDLN-----RLQOT---HCSHDSKNCRP-----DG-----NKPYSP--Y 463
QY 476 KDGKLLPQKSTIVQAGSYFNT-----FYFDAAL----- 506
Db 464 KSDRMIVESRNLFOAVFKAFDTAKIRHNLSINLGYDREKLSOLSHSDYVLQNAVOAYDL 523
QY 507 -----KKDIYRLNYSNTVG---YRFGGE-YT-----GYGSDDEFK 539
Db 524 ITPKPPPPNGSKNDYRVSIGTNTVTSPICRFGNNTYDTPRNIIGNGGYAAVODNV 583
QY 540 RAFGENSPYKKHCNRCGIYEPVLKYGKRRANNHSVS-----ISADFGDYFMPFASY 593
Db 584 R-----LGRWADVAGIRYDTRSHSEDSKSVSTGTHRNLSWNAAGVVLKPTWM 631
QY 594 SRTH-----RPMNIQEMYSQIGDSGVHTALKPERANTWQFG-----FNT 633
Db 632 DLTYRSTGRFLPSFAEMYGWRAGESLKTLDLKPESFNREAGIVFKGDFGNLEASYFNN 691
QY 634 YKGLLKQDDTLGLKLVGYSRIDN-----YIH-----NVYKG--WWDLNG 672
Db 692 AYRDLI-----AFGYETRTQNGQTSASGDPGYRNAQNAIAGINILGKIDWHGVWG 742
QY 673 DIPSWSVSSTGLAY-----TIQHRNFKDKVHKHGFELNLDY--GRFFTNL 716
Db 743 GLPDGLYST-LAYNRKVKDADIRADRTFTVSYLFDVAVQPSRYVLGLGYDHPDGIWGINT 801
QY 717 SYAYQKSTQPTNFSASESPNNASKEDQLKQGYCLSRVSPALPDYGRLEVGTWLGKNTL 776
Db 802 MFTYSK-----AKSVDEL-----LGSQAL 820
QY 777 LGGAMRYFGKSIRATAERYIDTNGGNTSNFRLGKRISIKOTETLARQPLIFDFYAAYE 836
Db 821 LNG-----NANAKAASRR-----TRPWYVTDVSGYIN 848
QY 837 PKKNLIFRAEVKNLFDORRYI 856
Db 849 IKKHLTLRAGVYNLLNRYV 868
```

RESULT 12

US-08-478-435-94

; Sequence 94, Application US/08478435

; Patent No. 592323

```
; GENERAL INFORMATION:
; APPLICANT: Loosmore, Sheena
; APPLICANT: Harkness, Robin
; APPLICANT: Schryvers, Anthony
; APPLICANT: Chong, Pele
; APPLICANT: Gray-Owen, Scott
; APPLICANT: Yang, Yan-ping
; APPLICANT: Murdin, Andrew
; APPLICANT: Klein, Michel
; TITLE OF INVENTION: Transferrin Receptor Genes
; NUMBER OF SEQUENCES: 147
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Sim & McBurney
; STREET: Suite 701, 330 University Avenue
; CITY: Toronto
; STATE: Ontario
; COUNTRY: Canada
; ZIP: M5G 1R7
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,435
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/337,483
; FILING DATE: 08-NOV-1994
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/175,116
; FILING DATE: 29-DEC-1993
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/148,968
; FILING DATE: 08-NOV-1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Stewart, Michael I
; REGISTRATION NUMBER: 24,973
; REFERENCE/DOCKET NUMBER: 1038-462 MIS:vg
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (416) 595-1155
; TELEFAX: (416) 595-1163
; INFORMATION FOR SEQ ID NO: 94:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 908 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; US-08-478-435-94
```

Query Match 5.5%; Score 269; DB 2; Length 908;

Best Local Similarity 21.2%; Pred. No. 1.7e-13;

Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;

QY 5 FRKPICFYLMGYTLHYHYAEDAGRAGSEAOIQVLEDEHVHKAK--RVPKDKKVFDTARA 62

Db 7 FRELNLCSLMTALPV---YAENV--QAEQAQEKQLDTIQVAKAKKQKTRDRNEVTGLGKL 61

QY 63 VSTRQDIFKSSE-NLDNIVRSIPG-AFTQDK--SSGIYSLNIRGDSGFRVNTWVDGIT 118

Db 62 VKSSDLSKQVNLNLDLTRYDPIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGVS 117

QY 119 Q-TFYSTSDAG--RAGGSSQFGASVDSNFIAGLDVVYKGSFGSAGINSLAGSANLRTLG 175

Db 118 QIOSYTAQAALGCTRTAGSGGAINIEYENKAVEISKGSNSSEYNGALAGSVAFOTKT 177

QY 176 VDDVVOGNTTYGLLK-GLTGNTSTKGNMAAIGARKWLESASGVLYGHRSRVAQNY 234

Db 178 AADIIEGKGWGIQSKTAYSGKDHALTQSLALAGR---SGGAELIITYTKRGRETHAH 233





Db 464 KSDRMIEESRNLFOAVFKAFDTAKIRHNLINLGYDRFKSOLSHSDYILQNAVQAYDL 523  
QY 507 -----KKDIYRLNYSNTWVG-----YRFGGE-YT-----GYGSDDEFK 539  
Db 524 ITPKKPPFPNGSKDNPRYSIGTKTNTVNTSPICRFGNNTYDCTPRNIGGNGYAAVQDNV 583  
QY 540 RAGENSPTKKHCNRCGIIYEPVLKKYKKRANHSVS-----ISADPGDYFMPFASY 593  
Db 584 R-----LGRWADYGAGIRYDIRSTHSEDKSVSTGTHRNLSWNAVGVLPKPTWM 631  
QY 594 SRTH-----RMPNIOEMFYSQIDSGVHTALKPERANTWQFG-----FNT 633  
Db 632 DLYRSTGRFLSPFAMYCWGRAGESLKTLDLAPKSFNREAGIVFKGDFGNLEASYFNN 691  
QY 634 YKGLLKQDPTLGLKLVGYSRIDN-----YIY-----NYGK--WMDLNG 672  
Db 692 AYRDLI-----AFGYETRTONGTSASGDPGYRNAQRIAGINILGKIDWHGVWG 742  
QY 673 DIPSWYSSTGLAY-----TIOHRNFKDKVHKHGFELNLYDY--GRFFTNL 716  
Db 743 GLPDGLYST-LAYNRIRKVKDADIRADRTFTVTSYLFDAVQPSRYVLGLGYDHPDGIWINT 801  
QY 717 SYAYQKSTQPTNFSASESPNNASKEDOLKOGYGLSRVSALPRDYGRLEVGTRWLGNKLT 776  
Db 802 MFTYSK-----AKSVDEL-----LGSQAL 820  
QY 777 LGGAMRYFGKSIRATAERYIDGTNGNTSNFRQLGKRSIKOTETLARQPLIFDYAAYE 836  
Db 821 LNG-----NANAKKAASRR-----TRPWYVTDVSGYYN 848  
QY 837 PKNLIFRAEVKNLFDPRYI 856  
Db 849 IKKHLTLRAGVYNLLNRYV 868

RESULT 14

US-08-478-373-94  
; Sequence 94, Application US/08478373  
; Patent No. 5922841  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena  
; APPLICANT: Harkness, Robin  
; APPLICANT: Schryvers, Anthony  
; APPLICANT: Chong, Pele  
; APPLICANT: Gray-Owen, Scott  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Mordin, Andrew  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Transferrin Receptor Genes  
; NUMBER OF SEQUENCES: 147  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: Patentin Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/478,373  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/337,483  
; FILING DATE: 08-NOV-1994  
; CLASSIFICATION: 435  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/175,116  
; FILING DATE: 29-DEC-1993

; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/148,968  
; FILING DATE: 08-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-463 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:  
; LENGTH: 908 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-478-373-94  
  
Query Match 5.5%; Score 269; DB 2; Length 908;  
Best Local Similarity 21.2%; Pred. No. 1.7e-13;  
Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;  
  
QY 5 FRLKPICFYLMGYLHYHYAEDAGRAGSEAOIQVLEDVHVKAK--RVPKDKKVFTDARA 62  
Db 7 FRLNLCLSLMTALPV---YAENV--QAEQAQEKQLDTIQVKKAKKQKTRDNEVTGLGKL 61  
QY 63 VSTRQDIFKSE--NLNDIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGVRVMTVDGIT 118  
Db 62 VKSSDLSLKEQVLNIRDLTRYDPGIAVVEQGRGNSSG---YSING-MDKNRVSLTVDGVS 117  
QY 119 Q-TFYSTSTDAG--RAGGSSQFGASVDSNFIAGLDVYVYKFSFGSAGINSIAGSANLRTLG 175  
Db 118 QIQSYTAQAALGGTTRTAGSSGAINIEYENVKAVEISKNSNSSEYNGALAGSVAFTKTK 177  
QY 176 VDDVQGNNTYGLLLK--GLTGTNSTKGNMAATGARKWLESAGSVGLYGHSSRSVAQNY 234  
Db 178 AADLIGEGKQWGIQSTAYSGKDHALTQSLAGR-----SGGAELIYIKRKGRETIHAH 233  
QY 235 RVGGGQGHIGFNGAEYLERRKQ-----RYFVQEGAL-----KFNSD-SGKWER-- 276  
Db 234 KDAGKG--VQSFNRLVLEDEKKEGGSQYRYFIVEECHNGYAAACKNLKEDASVKDERKT 291  
QY 277 -----DLORQWQYKYP-----YKNNQELQKYECHDKSKRNPENLAPQ 314  
Db 292 VSTQDYTGSRNLLANPLEYGSQSWLFRPGWHLNDRHYGVAVLERTQQTFTD--RDMTVPA 349  
QY 315 YDITPID--PSSLK-----QQSAGNLFKLEY-DGVFNKYTAQFRDLNLT 355  
Db 350 Y-FTSEDYVPGSLKGLGKYSKDNKAERLFOGEGSTLQIGYGTGVF-----YDERHTK 402  
QY 356 IGSRKIINRNYPNYGLSLNPNYTNLNTAAVNSGRQKYPKGSKFTGWGLLKDFETYNNAK 415  
Db 403 -----NR-----YGVVEY-VYHNADKDTWADYARLSYDRQG----- 431  
QY 416 ILDLNNTATPRLPRETELQTLGTFNHYFNHGYGNRPFEEELGDFGDDQDNGLYSYLGRF 475  
Db 432 -IDLN-----RLQOT---HCHSDGSKNCRP-----DG-----NKPYSF---Y 463  
QY 476 KDGKLLLPKQSTIVQAPAGSOYFNT-----FYFDAAL----- 506  
Db 464 KSDRMIEESRNLFOAVFKAFDTAKIRHNLINLGYDRFKSOLSHSDYILQNAVQAYDL 523  
QY 507 -----KKDIYRLNYSNTWVG-----YRFGGE-YT-----GYGSDDEFK 539  
Db 524 ITPKKPPFPNGSKDNPRYSIGTKTNTVNTSPICRFGNNTYDCTPRNIGGNGYAAVQDNV 583  
QY 540 RAGENSPTKKHCNRCGIIYEPVLKKYKKRANHSVS-----ISADPGDYFMPFASY 593  
Db 584 R-----LGRWADYGAGIRYDIRSTHSEDKSVSTGTHRNLSWNAVGVLPKPTWM 631  
QY 594 SRTH-----RMPNIOEMFYSQIDSGVHTALKPERANTWQFG-----FNT 633

Db 632 DLTYRSTGRLPSFAEMYGWRAGESLTKLTLKPEKSFNREAGIVFKGDFGNLEASYFNN 691  
QY 634 YKGLLKQDDTLGLKLGYRSRIDN-----YIH-----NVYK--WWDLNG 672  
Db 692 AYRDLI-----AFGYETRTQNGTSSASGDPGYRNAQNARIAGINILGKIDWHGCVG 742  
QY 673 DIPSWSVSTGLAY-----TIQHRNFKDKVHKHKGFELELNVDY--GRFPTNL 716  
Db 743 GLPDGLYST-LAYNRKVKDADIRADRTFTVSYLFDVAVQPSRVVLGLGYDHPDGWINGINT 801  
QY 717 SYAYQKSTQPTNFSDAESPNNASKEDQLKQGYCLRSVLSALPRDYGRLEVGTRWLNKLT 776  
Db 802 MFTYSK-----AKSVDEL-----LGSQAL 820  
QY 777 LGAMRYFGKSI RATAEERYIDGTNGTNTSNFRGLGRKRSIKQTETLARQPLIFDYAAYE 836  
Db 821 LNG-----NANAKAASRR-----TRPWYVTDVSGYYN 848  
QY 837 PKKNLIFRAEVKNLFDRIYI 856  
Db 849 IKKHLTLRAGVYNLLNRYV 868

RESULT 15  
US-08-474-671-94  
; Sequence 94, Application US/08474671  
; Patent No. 6008326  
; GENERAL INFORMATION:  
; APPLICANT: Loosmore, Sheena  
; APPLICANT: Harkness, Robin  
; APPLICANT: Schryvers, Anthony  
; APPLICANT: Chong, Pele  
; APPLICANT: Gray-Owen, Scott  
; APPLICANT: Yang, Yan-Ping  
; APPLICANT: Mordin, Andrew  
; APPLICANT: Klein, Michel  
; TITLE OF INVENTION: Transferrin Receptor Genes  
; NUMBER OF SEQUENCES: 147  
; CORRESPONDENCE ADDRESS:  
; ADDRESSEE: Sim & McBurney  
; STREET: Suite 701, 330 University Avenue  
; CITY: Toronto  
; STATE: Ontario  
; COUNTRY: Canada  
; ZIP: M5G 1R7  
; COMPUTER READABLE FORM:  
; MEDIUM TYPE: Floppy disk  
; COMPUTER: IBM PC compatible  
; OPERATING SYSTEM: PC-DOS/MS-DOS  
; SOFTWARE: PatentIn Release #1.0, Version #1.25  
; CURRENT APPLICATION DATA:  
; APPLICATION NUMBER: US/08/474,671  
; FILING DATE: 07-JUN-1995  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/337,483  
; FILING DATE: 08-NOV-1994  
; CLASSIFICATION: 530  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/175,116  
; FILING DATE: 29-DEC-1993  
; PRIOR APPLICATION DATA:  
; APPLICATION NUMBER: US 08/148,968  
; FILING DATE: 08-NOV-1993  
; ATTORNEY/AGENT INFORMATION:  
; NAME: Stewart, Michael I  
; REGISTRATION NUMBER: 24,973  
; REFERENCE/DOCKET NUMBER: 1038-465 MIS:vg  
; TELECOMMUNICATION INFORMATION:  
; TELEPHONE: (416) 595-1155  
; TELEFAX: (416) 595-1163  
; INFORMATION FOR SEQ ID NO: 94:  
; SEQUENCE CHARACTERISTICS:

; LENGTH: 908 amino acids  
; TYPE: amino acid  
; STRANDEDNESS: single  
; TOPOLOGY: linear  
; US-08-474-671-94

Query Match 5.5%; Score 269; DB 3; Length 908;  
Best Local Similarity 21.2%; Pred. No. 1.7e-13;  
Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;

QY 5 FRUKPICFYLMGVTLVHYSTAEADAGRAGSEAQIOVLVDVHVAK--RVPKDKKVFDDARA 62  
Db 7 FRNLILCLSLMTALPV---YAENV--QAEQAQEKOLDTIQVAKAKQKTRKREDNEVTGLGKL 61  
QY 63 VSTRQDIFKSS--NLNINIVRSIPG-AFTODK--SSGIYSLNIRGDSGFGFVRNMTVDGIT 118  
Db 62 VKSSDTLSKEQVLNIRDLTRYDPIAVVEOGRGASSG---YSIRG-MDKNRVSLTVDGVS 117  
QY 119 Q-TFYSTSTDAG--RAGGSSQFGASVDSNFIAGLDVVVKFSFGSAGINSLAGSANLRTL 175  
Db 118 QIQSYTAQAALGCTRTAGSSGAINIEYENVKAVEISKGSNSSEYNGALAGSVAFTQTKT 177  
QY 176 VDDVVOGNNYGLLLK-GLTGTNSTKGNMAAIGARKWLESAGSVGLYGHSSRRSVAQNY 234  
Db 178 AADIIGEGKQWGIQSKTAYSGKDHALTQSLALAGR---SGGAELLIYTKRRGREITHAH 233  
QY 235 RVGGGGHIGNFGAEYLERKQ-----RYFVOEGAL-----KFNSD-SGKWER-- 276  
Db 234 KDAGK--VOSEFNLVDEDKKEGSGQYRIFYVEEHCNGNYAACNKKLKEDASVOKERKT 291  
QY 277 -----DLQRQWKYKP-----YKNYNNQELQYIEGHDKSWRENLAPO 314  
Db 292 VSTQDYTGSNRLLANPLEYGSQSWLFRPGWHLNDRHYVGAVLERTQQTFT--RDMTVPA 349  
QY 315 YDITPID--PSSLK-----QQSAGNLFKLEY--DGVFNKYTAQFRDLNNTK 355  
Db 350 Y-FTSEDYVPGSLKGLGKYSGDNKABRLFVQEGEGSTLQIGIGYGTGVF-----YDERHTK 402  
QY 356 IGSRKLIINRYQFNGLSLNPTNLTAAYNSGRQKPKGSKFTGWLGLKDFETYNNAK 415  
Db 403 -----NR-----YGVY--VYHNADKDWADYARLSYDRQ----- 431  
QY 416 ILDNNTATRLPRETELQTLGTFNHYKKNRPEELGLFFDQDQNGLYSYLGRF 475  
Db 432 -IDLN-----RLQOT--HCSHDGSDKNCRP-----DG-----NKPYSE--Y 463  
QY 476 KDGKLLPKSTIVQAGSQYFNT-----FYFDAAL----- 506  
Db 464 KSDRMIEESRNLFOAVFKKAFDTAKIRHNLSINLGYDRFKSQLSHSDYVQLNAVOAYDL 523  
QY 507 -----KKDIYRLNYSNTVG---YRFGE--YT-----GYYSDDDEFK 539  
Db 524 ITPKPPFPNGSKDNPIRVSIGTKTNTVTPICRFGNTTYDCTPRNIGNGNYFAAODNV 583  
QY 540 RAFGENSPYKKHCNRSCTIYEVPLKKYKKRANNHSV-----ISADFGDYFMPFASY 593  
Db 584 R-----LGRWADVAGIRYDYRSTHSDSKSVSTGTHRNLSWAGVVLKPFPTWM 631  
QY 594 SRTH-----RMPNIQEMYFSQIGDSGVHTALKPERANTWQFG-----FNT 633  
Db 632 DLTYRSTGRLPSFAEMYGWRAGESLTKLTLKPEKSFNREAGIVFKGDFGNLEASYFNN 691  
QY 634 YKGLLKQDDTLGLKLGYRSRIDN-----YIH-----NVYK--WWDLNG 672  
Db 692 AYRDLI-----AFGYETRTQNGTSSASGDPGYRNAQNARIAGINILGKIDWHGCVG 742  
QY 673 DIPSWSVSTGLAY-----TIQHRNFKDKVHKHKGFELELNVDY--GRFPTNL 716  
Db 743 GLPDGLYST-LAYNRKVKDADIRADRTFTVSYLFDVAVQPSRVVLGLGYDHPDGWINGINT 801  
QY 717 SYAYQKSTQPTNFSDAESPNNASKEDQLKQGYCLRSVLSALPRDYGRLEVGTRWLNKLT 776

Db	802	MFTYSK-----AKSVDEL-----LGSQAL	820
Qy	777	LGAMRYFKSRATAERYIDGTNGGNTSNPQLGKRSIKOTETLARQPIIFTFYAAE	836
Db	821	LNG-----NANAKKAASRR-----TRPWYVTDVSGYIN	848
Qy	837	PKKNLIFRAEYKNNLFDRRYI	856
Db	849	IKKHLTRAGVYNLLNRYV	868

Search completed: July 24, 2002, 08:53:59  
Job time: 531 sec



Result No.	Score	Query Match	Length	DB ID	Description
1	4894	100.0	921	21	US-09-762-926-6
2	4779.5	97.7	922	21	US-09-762-926-4
3	4769.5	97.5	922	21	US-09-762-926-2
4	4707.5	96.2	922	17	US-09-303-518D-884
5	4703	96.1	888	17	US-09-303-518D-878
6	4696.5	96.0	922	17	US-09-303-518D-882
7	4348.5	88.9	871	17	US-09-303-518D-880

	Query Match	100.0%;	Score 4894;	DB 21;	Length 921;
	Best Local Similarity	100.0%;	Pred. No. 0;		
	Matches 921;	Conservative	0;	Mismatches	0;
				Indels	0;
QY	1	MRSSFRLKPCIFLMGVTLHYHSYAEDAGRAGSEAIQVLELVHVKAKRVPKDKKVF	TA	60	
Db	1	MRSSFRLKPCIFLMGVTLHYHSYAEDAGRAGSEAIQVLELVHVKAKRVPKDKKVF	TA	60	

## ALIGNMENTS

Qy	61	RAVSTRQDIFKSSNNLNI	VRSPGATQODKSSGIVSLNIRGDSGFRVNTWMDGITOT	120
Db	61	RAVSTRQDIFKSSNNLNI	VRSPGATQODKSSGIVSLNIRGDSGFRVNTWMDGITOT	120
Qy	121	FYSTSTADGRAGGSSQF	GSVDNSFTAGLDVWKGFSFGSAGINSLAGSANLRTLGVDDVV	180
Db	121	FYSTSTADGRAGGSSQF	GSVDNSFTAGLDVWKGFSFGSAGINSLAGSANLRTLGVDDVV	180
Qy	181	QGNNTYGLLLKGLTGT	NSTKGNMAAATGARKWLESGASGVLYGHSSRRSVAQNYRVGGG	240
Db	181	QGNNTYGLLLKGLTGT	NSTKGNMAAATGARKWLESGASGVLYGHSSRRSVAQNYRVGGG	240
Qy	241	QHIGNFAEYLERKQRY	FVQEGALFNDSGKWERDLQROQWKYPKYKNYNNQLOKYY	300
Db	241	QHIGNFAEYLERKQRY	FVQEGALFNDSGKWERDLQROQWKYPKYKNYNNQLOKYY	300
Qy	301	EGHDKSWRENLAPOYD	ITPIDPSLKKQSSAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRK	360
Db	301	EGHDKSWRENLAPOYD	ITPIDPSLKKQSSAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRK	360
Qy	361	IINRNYOFNYGLSLNP	YTNLNLTAAYNSGRQKYPKGSKGTGWGLLKDFETYNNAKILDLN	420
Db	361	IINRNYOFNYGLSLNP	YTNLNLTAAYNSGRQKYPKGSKGTGWGLLKDFETYNNAKILDLN	420
Qy	421	NTATFRLPRETELOTT	ILGPNYFHNEYCKNRFPEELGLFDPGDDQDNGLYSLGRFKPGDKG	480
Db	421	NTATFRLPRETELOTT	ILGPNYFHNEYCKNRFPEELGLFDPGDDQDNGLYSLGRFKPGDKG	480
Qy	481	LLPOKSTIIVOPAGSO	QFNTFYFPAALKKDIIYRLNYNTVGRFGEYTYGYSDDDEPKR	540
Db	481	LLPOKSTIIVOPAGSO	QFNTFYFPAALKKDIIYRLNYNTVGRFGEYTYGYSDDDEPKR	540
Qy	541	AFGNSPTYKKHCNR	SGCIEPVLYKKYKGRANNHSSVSIADFGDYFMPFASYSTRHP	600
Db	541	AFGNSPTYKKHCNR	SGCIEPVLYKKYKGRANNHSSVSIADFGDYFMPFASYSTRHP	600
Qy	601	NIQEMYSQIGDSGV	HTALPERANTWQFGNTYKKGLLKQDDTLGLKLGVYRSRIDNYI	660
Db	601	NIQEMYSQIGDSGV	HTALPERANTWQFGNTYKKGLLKQDDTLGLKLGVYRSRIDNYI	660
Qy	661	HNVYKQWMDLNGDI	PSWVSSTGLAYTIOHRNPKDKVHKHGFLELNYDYGREFTNL	720
Db	661	HNVYKQWMDLNGDI	PSWVSSTGLAYTIOHRNPKDKVHKHGFLELNYDYGREFTNL	720
Qy	721	QKSTQPTNFDASES	PNNASKEDOLKQGYGLRSVALPRDYGRLEVGTFRWLGNKLT	780
Db	721	QKSTQPTNFDASES	PNNASKEDOLKQGYGLRSVALPRDYGRLEVGTFRWLGNKLT	780
Qy	781	MRVFGKSIRATAERY	IDGTNGTNTSNFQJGKRSIKQETLARQPLDFDYAAEYEPKKN	840
Db	781	MRVFGKSIRATAERY	IDGTNGTNTSNFQJGKRSIKQETLARQPLDFDYAAEYEPKKN	840
Qy	841	LIFRAEYKVLFD	RRYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTL	900
Db	841	LIFRAEYKVLFD	RRYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTL	900
Qy	901	KSVLTNFARGTFL	ITMYSYKF 921	
Db	901	KSVLTNFARGTFL	ITMYSYKF 921	

## RESULT 2

[illegible]

QY 900 SKSVLTNFARGRTFLITMSYKF 921  
Db 901 SKSVLTNFARGRTFLITMSYKF 922  
RESULT 3  
US-09-762-926-2  
; Sequence 2, Application US/09762926  
; GENERAL INFORMATION:  
; APPLICANT: Thonnard, Joelle  
; FILE REFERENCE: BM45330  
; CURRENT APPLICATION NUMBER: US/09/762,926  
; CURRENT FILING DATE: 2001-02-14  
; PRIOR APPLICATION NUMBER: PCT/JP99/05989  
; PRIOR FILING DATE: 1999-08-13  
; PRIOR APPLICATION NUMBER: GB 9818004.5  
; PRIOR FILING DATE: 1998-08-18  
; NUMBER OF SEQ ID NOS: 9  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 2  
; LENGTH: 922  
; TYPE: PRT  
; ORGANISM: Bacteria  
US-09-762-926-2

Query Match 97.5%; Score 4769.5; DB 21; Length 922;  
Best Local Similarity 97.7%; Pred. No. 0;  
Matches 901; Conservative 8; Mismatches 12; Indels 1; Gaps 1;  
QY 1 MRSSFLKPCFCYLMGVTLHYHYAEDAGRAGSEAOIQVLEDDVHVAKRVPKDKKVFDA 60  
Db 1 MRSSFLKPCFCYLMGVTLHYHYAEDAGRAGSEAOIQVLEDDVHVAKRVPKDKKVFDA 60  
QY 61 RAVSTRODIFKSSNLDNIIVRSIPGAFTQDDKSSGIVSLNIRGDSGFRVNTWVDGITOT 120  
Db 61 RAVSTRODIFKSSNLDNIIVRSIPGAFTQDDKSSGIVSLNIRGDSGFRVNTWVDGITOT 120  
QY 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV 180  
Db 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV 180  
QY 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGG 240  
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGG 240  
QY 241 OHICNFGAEYLERRKQRYFVQEGALKFNSDGKWERDLRQOQWKYKPYKNYN-OELQY 299  
Db 241 OHICNFGAEYLERRKQRYFVQEGALKFNSDGKWERDLRQOQWKYKPYKNYN-OELQY 299  
QY 300 IEHDKSWRENLAPOYDITPIDPSSLKQOQAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 359  
Db 300 IEHDKSWRENLAPOYDITPIDPSSLKQOQAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 359  
QY 360 KIINRNYQFNYGLSLNPTNLNTAAYNSGRQKYPKGSKFTGWLKDFETYNNAKILDL 419  
Db 360 KIINRNYQFNYGLSLNPTNLNTAAYNSGRQKYPKGSKFTGWLKDFETYNNAKILDL 419  
QY 420 NNTATPLPRETELQTTGLGNYFNHNEYGKRPPEELGLFFDGDQDNGLYSYLGRFGDK 479  
Db 420 NNTATPLPRETELQTTGLGNYFNHNEYGKRPPEELGLFFDGDQDNGLYSYLGRFGDK 479  
QY 480 GLLPQKSTIVQAGSQYFNTFYDAALKDIIYRLNLTNTVYRFGGEYTYGYSDDDEFK 539  
Db 480 GLLPQKSTIVQAGSQYFNTFYDAALKDIIYRLNLTNTVYRFGGEYTYGYSDDDEFK 539  
QY 540 RAFGENSPYKHKCNKSCGIEPVLYKYGKRRANNHVSISADFGDYFMPFASYSRTHRM 599  
Db 540 RAFGENSPYKHKCNKSCGIEPVLYKYGKRRANNHVSISADFGDYFMPFASYSRTHRM 599  
QY 600 PNQEMVFSQIGDSGVITALKPERANTWQGFNTYKGLLKQDDTLGLKLVGYRSRDNY 659

Db 601 PNQEMVFSQIGDSGVITALKPERANTWQGFNTYKGLLKQDDTLGLKLVGYRSRDNY 660  
QY 660 IHNVYGKWDNLNGDIPSWVSSTGLAYTIQHRNFKDKVHKHGFLELNYDYGRFFTNLSYA 719  
Db 661 IHNVYGKWDNLNGDIPSWVSSTGLAYTIQHRNFKDKVHKHGFLELNYDYGRFFTNLSYA 720  
QY 720 YOKSTQPTNFSADASESPNNASKEDQLKQGYGLSRVSALPRDYGRELVGTWGLGNKLTGG 779  
Db 721 YOKSTQPTNFSADASESPNNASKEDQLKQGYGLSRVSALPRDYGRELVGTWGLGNKLTGG 780  
QY 780 AMRYFGKSIIRATAERYIDGTNGGNTSNFROLGKRSIKQETTLARQPLIFDYAAYEPKK 839  
Db 781 AMRYFGKSIIRATAERYIDGTNGGNTSNFROLGKRSIKQETTLARQPLIFDYAAYEPKK 840  
QY 840 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSDFDPKDKDEDTVCNADKTLCKNGYGGT 899  
Db 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSDFDPKDKDEDTVCNADKTLCKNGYGGT 900  
QY 900 SKSVLTNFARGRTFLITMSYKF 921  
Db 901 SKSVLTNFARGRTFLITMSYKF 922

## RESULT 4

US-09-303-518D-884  
; Sequence 884, Application US/09303518D  
; GENERAL INFORMATION:  
; APPLICANT: Scarlato, Vincenzo  
; APPLICANT: Masignani, Vega  
; APPLICANT: Rappuoli, Rino  
; APPLICANT: Pizza, Mariagrazia  
; TITLE OF INVENTION: Neisserial Antigens  
; FILE REFERENCE: CHIR0160  
; CURRENT APPLICATION NUMBER: US/09/303,518D  
; CURRENT FILING DATE: 1999-04-30  
; NUMBER OF SEQ ID NOS: 1098  
; SOFTWARE: PatentIn version 3.1  
; SEQ ID NO 884  
; LENGTH: 922  
; TYPE: PRT  
; ORGANISM: Neisseria gonorrhoeae  
US-09-303-518D-884

Query Match 96.2%; Score 4707.5; DB 17; Length 922;  
Best Local Similarity 96.0%; Pred. No. 0;  
Matches 885; Conservative 15; Mismatches 21; Indels 1; Gaps 1;

QY 1 MRSSFLKPCFCYLMGVTLHYHYAEDAGRAGSEAOIQVLEDDVHVAKRVPKDKKVFDA 60  
Db 1 MRSSFLKPCFCYLMGVTLHYHYAEDAGRAGSEAOIQVLEDDVHVAKRVPKDKKVFDA 60  
QY 61 RAVSTRODIFKSSNLDNIIVRSIPGAFTQDDKSSGIVSLNIRGDSGFRVNTWVDGITOT 120  
Db 61 RAVSTRODIFKSSNLDNIIVRSIPGAFTQDDKSSGIVSLNIRGDSGFRVNTWVDGITOT 120  
QY 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV 180  
Db 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV 180  
QY 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGG 240  
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGG 240  
QY 241 OHICNFGAEYLERRKQRYFVQEGALKFNSDGKWERDLRQOQWKYKPYKNYN-OELQY 299  
Db 241 OHICNFGAEYLERRKQRYFVQEGALKFNSDGKWERDLRQOQWKYKPYKNYN-OELQY 299  
QY 300 IEHDKSWRENLAPOYDITPIDPSSLKQOQAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 359  
Db 301 IEHDKSWRENLAPOYDITPIDPSSLKQOQAGNLFKLEYDGVFNKYTAQFRDLNLTIGSR 360





7  
PFCIII.T

; GENERAL INFORMATION:

```

1 APPLICANT: Pizzia, Maria Grazia
2
3 TITLE OF INVENTION: Neisserial Antigens
4
5 FILE REFERENCE: CHIR0160
6
7 CURRENT APPLICATION NUMBER: US/09/303,5
8
9 CURRENT FILING DATE: 1999-04-30
10
11 NUMBER OF SEQ ID NOS: 1098
12
13 SOFTWARE: PatentIn version 3.1
14
15 SEQ ID NO 880
16
17 LENGTH: 871
18
19 TYPE: PRT
20
21 ORGANISM: Neisseria meningitidis
22
23 FEATURE:
24
25 NAME/KEY: misc_feature
26
27 LOCATION: (22)..(22)
28
29 OTHER INFORMATION: Xaa= any amino acid
30
31 NAME/KEY: misc_feature
32
33 LOCATION: (31)..(31)
34
35 OTHER INFORMATION: Xaa= any amino acid
36
37 NAME/KEY: misc_feature
38
39 LOCATION: (38)..(38)
40
41 OTHER INFORMATION: Xaa= any amino acid
42
43 NAME/KEY: misc_feature
44
45 LOCATION: (40)..(40)
46
47 OTHER INFORMATION: Xaa= any amino acid
48
49 NAME/KEY: misc_feature
50
51 LOCATION: (45)..(45)
52
53 OTHER INFORMATION: Xaa= any amino acid
54
55 NAME/KEY: misc_feature
56
57 LOCATION: (52)..(52)
58
59 OTHER INFORMATION: Xaa= any amino acid
60
61 NAME/KEY: misc_feature
62
63 LOCATION: (68)..(68)
64
65 OTHER INFORMATION: Xaa= any amino acid
66
67 NAME/KEY: misc_feature
68
69 LOCATION: (95)..(95)
70
71 OTHER INFORMATION: Xaa= any amino acid
72
73 NAME/KEY: misc_feature
74
75 LOCATION: (124)..(124)
76
77 OTHER INFORMATION: Xaa= any amino acid
78
79 NAME/KEY: misc_feature
80
81 LOCATION: (133)..(133)
82
83 OTHER INFORMATION: Xaa= any amino acid
84
85 NAME/KEY: misc_feature
86
87 LOCATION: (277)..(277)
88
89 OTHER INFORMATION: Xaa= any amino acid
90
91 NAME/KEY: misc_feature
92
93 LOCATION: (355)..(355)
94
95 OTHER INFORMATION: Xaa= any amino acid
96
97 NAME/KEY: misc_feature
98
99 LOCATION: (370)..(370)
100
101 OTHER INFORMATION: Xaa= any amino acid
102
103 NAME/KEY: misc_feature
104
105 LOCATION: (414)..(414)
106
107 OTHER INFORMATION: Xaa= any amino acid
108
109 NAME/KEY: misc_feature
110
111 LOCATION: (477)..(477)
112
113 OTHER INFORMATION: Xaa= any amino acid
114
115 NAME/KEY: misc_feature
116
117 LOCATION: (483)..(483)
118
119 OTHER INFORMATION: Xaa= any amino acid

```

; OTHER INFORMATION: Xaa= any amino acid

```

; NAME/KEY: misc_feature
; LOCATION: (500)..(500)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (608)..(608)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (660)..(660)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (749)..(749)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (753)..(754)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (767)..(767)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (781)..(781)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (790)..(790)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (861)..(861)
; OTHER INFORMATION: Xaa= any amino acid
US-09-303-518D-880

```

Query Match	88.9%;	Score 4348.5;	DB 17;	Length 871;
Best Local Similarity	94.6%;	Pred. No. 0;		
Matches 824;	Conservative	8;	Mismatches 38;	Indels 1;
Gaps				
QY	52	KDKVFTDARAVSTRODIPKFSSENLDNIVRSIPCAFTQODKSSGIVSLNIRGDSGGRVYV	111	
Db	1	KDKVFTDARAVSTRODIPKFSSENLDNIVRXIPCAFTXQKSSGXVSLNIRXDSGGRVYV	60	
QY	112	TWVDGITQTFYSTSDAGRAGGSSQFCASVDSNFIAGLDVVKGFSFGSAGINSLAGSAML	171	
Db	61	TWVDGITXTFYSTSDAGRAGGSSQFCASVDSNFXAGLDVVGFSFGSAGINSLAGSAML	120	
QY	172	RTLGVDDVDVGGNTYGLLLKGLTGNTSTKGNMAAIGARKWLESASGVLYCHSPRSVA	231	
Db	121	RTLGVDDVDVGGNTYGLLLKGLTGNTSTKGNMAAIGARKWLESASGVLYCHSPRSVA	180	
QY	232	QNYRVGGGGHIGFAGAEYLERRKQRYFVQEGALKFNSDSGKWERDLQCOOKYKPYKNY	291	
Db	181	QNYRVGGGGHIGFAGAEYLERRKQRYFQEGGLKFNNSGKWERDFQKSYWKTKWQKY	240	
QY	292	N-NOELQKYITEGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQPR	350	
Db	241	DAPOELQKYITEGHDKSWRENLAPOYDITPIDPSSLKXQOSAGNLFKLEYDGVFNKYTAQPR	300	
QY	351	DLNATKIGSRKIIIRNRYQFNVGLSLNPYTNLNLTAAYNSGRQKYPKSGKTGCLLKDFET	410	
Db	301	DLNATKIGSRKIIIRNRYQFNVGLSLNPYTNLNLTAAYNSGRQKYPKSGKTGCLLKDFET	360	
QY	411	YNNAKILDNLANTAFRLPRETELQTTLGFNYPHNEYKGNRFPDEELGLFDGPDQDNGLYS	470	
Db	361	YNNAKILDNLANTAFRLPRETELQTTLGFNYPHNEYKGNRFPDEELGLFDGPDQDNGLYS	420	
QY	471	YLGRFKDGKLLPKOKSTIVOPAGSQYFNTFYDAAULKDIYRLNLTSTNVTGVYRFGGEYTG	530	
Db	421	YLGRFKDGKLLPKOKSTIVOPAGSQYFNTFYDAAULKDIYRLNLTSTNVTGVYRFGGEYTG	480	
QY	531	YYGSDDEFKRAFGENSPYKHKHCNRCSGIYEPVLKVKYKKRNNHNSVSIADFGDYFMPF	590	
Db	481	YYXSDDEFKRAFGENSPYKXHCNRCSGIYEPVLKVKYKKRNNHNSVSIADFGDYFMPF	540	
QY	591	ASYSRTHRMPNIOEMFYQSIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKLIV	650	
Db	541	ASYSRTHRMPNIOEMFYQSIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKLIV	600	

Qy	651	GYRSRIDNYIHNVYGVKWDLNGDIPSVSWSTGLAYTIQHNRNFKDKVHKHGFELNLYDYG	7110
Db	601	GYRSRIDXYIHNVYGVKWDLNGNIPSVSWSTGLAYTIQHNRNFKDKVHKHGFELNLYDX	660
Qy	711	REFTNLSYAVOKSTQPTNFSDAESPNNASKEQOLKOGYGLSRVSALPRDYGLEVGTRW	770
Db	661	REFTNLSYAVOKSQPTNFSDAESPNNASKEQOLKOGYGLSRVSALPRDYGREVGTRW	720
Qy	771	LGNNKLTGGGAMRYEGKISIRATAEERYIDGTNGGNTSNFROLGKRISIKQETTLARQPLIFD	830
Db	721	LGNNKLTGGGAMRYEGKISIRATAEERYIDXTNGXNTSNFROLGKRISIXQETTLARQPLIFD	780
Qy	831	FYAAEYEPKKNLIFRAEYKVNLFDRRYIDPLDAGNDAATQRYYSFDPKDKDEEDVTCNADKT	890
Db	781	XYAAEYEPKKNLIFRAEYKVNLFDRRYIDPLDAGNDAATQRYYSFDPKDKDEEVTCTNDNT	840
Qy	891	LCNGKYGGTGSKSVLTNFARGRTFLITMSYKF	921
Db	841	LCNGKYGGTGSKSVLTNFARGXTFLITMSYKF	871

RESULT 8  
 US-09-719-190-2  
 ; Sequence 2, Application US/09719190  
 ; GENERAL INFORMATION:  
 ; APPLICANT: Thonnard, Joelle  
 ; TITLE OF INVENTION: Moraxella Catharrahalis Polynucleotides  
 ; TYPE OF INVENTION: and Polypeptides  
 ; FILE REFERENCE: BM45326  
 ; CURRENT APPLICATION NUMBER: US/09/719,190  
 ; CURRENT FILING DATE: 2001-05-14  
 ; PRIOR APPLICATION NUMBER: PCT/EP99/03824  
 ; PRIOR FILING DATE: 1999-05-31  
 ; NUMBER OF SEQ ID NOS: 6  
 ; SOFTWARE: Fast-Seq for Windows Version 4.0  
 ; SEQ ID NO 2  
 ; LENGTH: 947  
 ; TYPE: PRT  
 ; ORGANISM: Moraxella catarrhalis  
 US-09-719-190-2

Query Match	52.9%;	Score 2588.5;	DB 21;	Length 947;
Best Local Similarity	55.6%;	Pred. NO. 4e-232;		
Matches 503;	Conservative 144;	Mismatches 232;	Indels .25;	Gaps 12;

39	Qy	VLEDVHVKAAR-VPKDKKVVTDARAVSTRODIPKSSSENLDNIVRSPGAFQTQDKSSGIV	97
40		: :	
41		: :	
42		: :	
43		: :	
44		: :	
45		: :	
46		: :	
47		: :	
48	Db	ILDEVVYATATNGTKKSQKPFKASATSVRENVFNAGSINDAIYRVPVPGAFQTQDKSSGIV	107
49		: :	
50		: :	
51		: :	
52		: :	
53		: :	
54		: :	
55		: :	
56		: :	
57		: :	
58		: :	
59		: :	
60		: :	
61		: :	
62		: :	
63		: :	
64		: :	
65		: :	
66		: :	
67		: :	
68		: :	
69		: :	
70		: :	
71		: :	
72		: :	
73		: :	
74		: :	
75		: :	
76		: :	
77		: :	
78		: :	
79		: :	
80		: :	
81		: :	
82		: :	
83		: :	
84		: :	
85		: :	
86		: :	
87		: :	
88		: :	
89		: :	
90		: :	
91		: :	
92		: :	
93		: :	
94		: :	
95		: :	
96		: :	
97		: :	
98	Qy	SLNTRGDSGGERVNTWVDGTTQTFYSTSDAGRAGGSSQFGASVDSNFTAGLDVVKGSFS	157
99		: :	
100		: :	
101		: :	
102		: :	
103		: :	
104		: :	
105		: :	
106		: :	
107		: :	
108	Db	SLNVRGDSGGERANSWVDGTTQTFYSTSDAGRGGTSQEGAVIDQNF TAGVELNKGSEN	167
109		: :	
110		: :	
111		: :	
112		: :	
113		: :	
114		: :	
115		: :	
116		: :	
117		: :	
118		: :	
119		: :	
120		: :	
121		: :	
122		: :	
123		: :	
124		: :	
125		: :	
126		: :	
127		: :	
128		: :	
129		: :	
130		: :	
131		: :	
132		: :	
133		: :	
134		: :	
135		: :	
136		: :	
137		: :	
138		: :	
139		: :	
140		: :	
141		: :	
142		: :	
143		: :	
144		: :	
145		: :	
146		: :	
147		: :	
148		: :	
149		: :	
150		: :	
151		: :	
152		: :	
153		: :	
154		: :	
155		: :	
156		: :	
157		: :	
158	Qy	GSAGINSLAGSANSRLTLGVDDVVQNNNTYGLLLKGLTGTNSTKGNMAWAIGARKWUESGA	217
159		: :	
160		: :	
161		: :	
162		: :	
163</			

```
Db 407 VLAHNVGQYKPKGSTFTGKLDKDFETKNTANLFDLNNHTFNLPKQMDLTTVGLNI 466
Qy 442 FHNYGKNRPEELGLFDDGDDONGLYSYL-GRFGDKGLLPQKSTIVPAGSQYNTFE 500
Db 467 LHNEYSKNRFDELGLFVYTNLLCGGFGYDAGCGGRFQGTSTLPPKKSIVIVQPSGKQRFHSI 526
Qy 501 YFDAALKKDIYRLNYSNTVGYRGEGVYTYGYSDDDEFKRAFGENSPTYKKHCNRSCTIY 560
Db 527 YLDTSLQDKYOLDYISVNASQYRFSGGEHASYSSQKFEQDKFGEDSQIYKHCSPSCDVY 586
Qy 561 EPLVTKGKKHAINHVSYLSAKYDTGEMFVSFARTHRMNPNIQEMFVSQICDGVNTALK 646
Db 621 PERANTWQFGENTYKGLLKQDDTLGLKLVGYRSRINYIHNHYGKWDNLNGDIPSWSS 680
Qy 647 PEQANTYOLGFEKRNLLTDNDFLGLKVVGYQSRINNYIHNHYGKWDYKNP-PSWVTS 705
Db 681 TGL-AYTIQHRNFKDKVHKHGFLELNDYGRFFTNLSYAYOKSTQPTNFSASESPNNA 739
Db 706 GALKGDTIOHNRWQMPVHKQGLELEINYDAGRYFTNLSYARQTDQPTNFSASESPNNS 765
Qy 740 SKEDQLQGGYGLSVSALPRDYGRLEVGRVWLGKNTLGGAMRYFGKSIIRATAEERYIDG 799
Db 766 SKEDQLTQGGYGLSVSMLPKDYGRFELGVGRFDDKLTIGSAVRYYGOSPRATIEPRVIDG 825
Qy 800 TNGGNTSNFRLGKRSIKQETELARQPLIFDFYAAEPKKNLIFRAEVKNLFDRLYIDPL 859
Db 826 THGGNTSHSDDKGAHVIKQIEMLKRPQLVHDFYVAYEPIKDLVNRDLVQNAFRLYIDPL 885
Qy 860 DAGNDAATQRYSSFDPKDKDEDVTCNADKTLCL-NGKYGTGTSKSVLTNFAARGRTFLITM 917
Db 886 DANNDATQRYHSY-INDADEGAPCAAGQ-LCKPDAKYGGTTRSVLTNFAKGRSLFSM 943
Qy 918 SYKF 921
Db 944 TYKW 947

RESULT 9
US-09-303-518D-876
; SEQUENCE 876, Application US/09303518D
; GENERAL INFORMATION:
; APPLICANT: Scarlato, Vincenzo
; APPLICANT: Masignani, Vega
; APPLICANT: Rappuoli, Rino
; APPLICANT: Pizza, Mariagrazia
; APPLICANT: Grandi, Guido
; TITLE OF INVENTION: Neisserial Antigens
; CURRENT APPLICATION NUMBER: US/09/303,518D
; CURRENT FILING DATE: 1999-04-30
; NUMBER OF SEQ ID NOS: 1098
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 876
; LENGTH: 393
; TYPE: PRP
; ORGANISM: Neisseria meningitidis
; FEATURE:
; NAME/KEY: misc_feature
; LOCATION: (21)..(21)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (104)..(104)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (163)..(163)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (165)..(165)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (171)..(178)
```

```
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (301)..(301)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (339)..(339)
; OTHER INFORMATION: Xaa= any amino acid
; NAME/KEY: misc_feature
; LOCATION: (353)..(353)
; OTHER INFORMATION: Xaa= any amino acid
US-09-303-518D-876

Query Match 40.8%; Score 1996; DB 17; Length 393;
Best Local Similarity 95.2%; Pred. No. 2.9e-177;
Matches 373; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

Qy 530 GYGSDDDEFKRAFGENSPTYKKHCNRSCTIYEPVLYKKYKKRANHSHSVISADFGDYFMP 589
Db 2 GYGSDDDEFKRAFGENSPTYKKHCNRSCTIYEPVLYKKYKKRANHSHSVISADFGDYFMP 61
Qy 590 FASYSTRHRMNPNIQEMFYSGIDSGVHTALKPERANTWQFGENTYKGLLKQDDTLGLKL 649
Db 62 FASYSTRHRMNPNIQEMFYSGIDSGVHTALKPERANTWQFGENTYKGLLKQDDTLGLKL 121
Qy 650 VGYRSRINDYIHNHYGKWDNLNGDIPSWVSTGLAYTIQHNRNFKDKVHKHGFLELNDY 709
Db 122 VGYRSRINDYIHNHYGKWDNLNGDIPSWVSTGLAYTIQHNRNFKDKVHKHGFLELNDY 181
Qy 710 GRFTNLSYAYOKSTQPTNFSASESPNNAKEDQLKQGLSVSALPRDYGRLEVGR 769
Db 182 GRFTNLSYAYOKSTQPTNFSASESPNNAKEDQLKQGLSVSALPRDYGRLEVGR 241
Qy 770 WLGKNTLGGAMRYFGKSIIRATAEERYIDGTNGGNTSNFRLGKRSIKQETELARQPLIF 829
Db 242 WLGKNTLGGAMRYFGKSIIRATAEERYIDGTNGGNTSNFRLGKRSIKQETELARQPLIX 301
Qy 830 DFYAAEPKKNLIFRAEVKNLFDRLYIDPLDAGNDAATQRYSSFDPKDKDEDVTCNADK 889
Db 302 DFNAAYEPKKNLIFRAEVKNLFDRLYIDPLDAGNDAAXERYISSFDPKDKDXDVTNADK 361
Qy 890 TLCNGKYGGTSKSVLTNFAARGRTFLITMSYKF 921
Db 362 TLCNGKYGGTSKSVLTNFAARGRTFLITMSYKF 393

RESULT 10
US-09-543-681A-8096
; SEQUENCE 8096, Application US/09543681A
; GENERAL INFORMATION:
; APPLICANT: GARY BRETON
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PROTEUS MIRA
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1002-001
; CURRENT APPLICATION NUMBER: US/09/543,681A
; CURRENT FILING DATE: 2000-04-05
; PRIOR APPLICATION NUMBER: US 60/128,706
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 8344
; SEQ ID NO 8096
; LENGTH: 760
; TYPE: PRP
; ORGANISM: Proteus mirabilis
US-09-543-681A-8096

Query Match 15.1%; Score 738.5; DB 19; Length 760;
Best Local Similarity 26.4%; Pred. No. 5.2e-59;
Matches 235; Conservative 133; Mismatches 324; Indels 199; Gaps 28;

Qy 47 AKRVPDKKIVTDARAVSTRODIFKSSLENLNIIVRSIPGAFTQODKSSGIVSLNIRGDSG 106
Db 53 AQOTPEQKALSKPGAYSATGEI-NNLSSVEQALRSTPGTTTQMDASQPGVGVNIRGLSG 111
```



Db 957 -----DV-----LAFTLGRGRTLQGTLEXPQF 977

RESULT 12

US-09-252-991A-27095  
; Sequence 27095, Application US/09252991A  
; GENERAL INFORMATION:  
; APPLICANT: Marc J. Rubenfield et al.  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO PSEUDOMONAS  
; FILE REFERENCE: 107196.136  
; CURRENT APPLICATION NUMBER: US/09/252,991A  
; CURRENT FILING DATE: 1999-02-18  
; PRIOR APPLICATION NUMBER: US 60/074,788  
; PRIOR FILING DATE: 1998-02-18  
; PRIOR APPLICATION NUMBER: US 60/094,190  
; PRIOR FILING DATE: 1998-07-27  
; NUMBER OF SEQ ID NOS: 33142  
; SEQ ID NO 27095  
; LENGTH: 992  
; TYPE: PRT  
; ORGANISM: Pseudomonas aeruginosa  
US-09-252-991A-27095

Query Match 8.7%; Score 426; DB 16; Length 992;  
Best Local Similarity 22.9%; Pred. No. 1.5e-29;  
Matches 235; Conservative 132; Mismatches 377; Indels 284; Gaps 44;

QY 22 YSAEADAGRAGSEAOIQV-----LEDVHVAKRV-PKDKKVF 58  
Db 121 YSVLEGLQRLVSLTGLQVAGTSGAYLLVEPRMPGETPADLSPVYSAELADPQRETYT 180  
QY 59 DARAVSTRQDIFKSENLD-----NIVRSIPGAFQODKSSGIVSLNIRGDSGFRV 110  
Db 181 -----APRSSVYLSSSIDRFRVSGVLDLQIGPGVQVDSRGGALDVNIRGQOSRV 235  
QY 111 NTWVDGITQFYSTSDTAGRAGSSQFASVDSNFAGLDVVGSGSAGINSAGSAN 170  
Db 236 AVRVDGAEQ-----ALDVRYGAGTQORSYIDPLVSVTVTDKGPSTRGAI---GGSVE 287  
QY 171 LRLTGVDVVQGNNTYGLLLKLTGNTSGNMAAATGARKWLESGASV-----GVLYGH 225  
Db 288 MRTIGVKDILVDGKDLGVRGTGDVWNN--GVAPQHSASCKTENLSSVPHDGRGSLFGS 344  
QY 226 SRR--SVAQNYRVGGGQHIGNFGAELERRKQRYFVQEGALKFNDSGKWERDLQKQW 283  
Db 345 QAKSGSAFAFYR---NEHL-DLVAAYAOQNOGYE-----SGKKGD----- 382  
QY 284 KYPKYKNYNOE---LQKYIEGHD-----KSWRENLAPOYDIT----- 318  
Db 383 RYRYNRYGRESSVAKVYNAGBEVLNSSSETESYLLKATWR--IADEHTLDLGYRYYDG 440  
QY 319 ---PIDPSSLKQOSAGNLFLEYDGV-FNKYTAQFRLNTKIGSRKIINRYQFNGLSL 374  
Db 441 RTGEIMPDIIRGTAGIYQYPLSEWKIDITYTARYLPEN-----NPLVDLSTGLMW 493  
QY 375 NPYTNLNLTAAYNSGROKYPKSGKFTGWLLKDFETYNNNAKI-LDLNNTATFRL----- 427  
Db 494 TDAKSDMLTVLAPRSQAYSRRNWT-----RODNERIGDNLNVARFETDFGDFK 544  
QY 428 -----PRETELOTTIGFN-----YFNEYCKN-----RPFEEGLGFLFDG- 461  
Db 545 LDLGGSFQVEDIQPKQSVVTTLDINANRTLDRATROEYGLNGKLEFPKPERLTLWGGR 604  
QY 462 ----PQDNGLYSLGRFGKDKLLPKQSTIVOPA--GSOYF---NTFFYDAA---LKK 508  
Db 605 YSHFNKDNIGISAPREDMDRFI-----TVSRPGYGYMMWFPDQNGQYTDATDPLNN 660  
QY 509 DIYRLNYSNTVGYRFGGEYTYGSDPEFKRAFGENSPYKKHCNRCSCGIYEPV-LKKY 567  
Db 661 GIVNTNNTNPFEGIPF-----DEFGPA-----NVTVHPSRVTVNTVGYNSKK 703

QY 568 GKRRANNHVSISADP---CDYFMPFPASYSRTHMPNIOEMYSQIGDSGVHTA--LKPE 622  
Db 704 GSSRGGGFSPAFGINFELAPDFTFV-YASYTEGLRLPSLFE---TSQGTLOQPEPKDLKPE 759  
QY 623 RANTWQFGFNTYKKGLLKQDDTLGLKLVGYRSRIDNVIHNVYKWKWDLNGLDIPSWSSG 682  
Db 760 RSRWEIGASALRDSLLADGDSAAIKLAYFNNTKNTYTRYD-----PQMGMLMT 810  
QY 683 LAVTIOHNFKDKVHKHGFLELNVDYGRFTNLSYAYQKSTQPTNFSASESPNNASKE 742  
Db 811 FSNT-----DSYRTSGLEQSHYDAGRFAVADLSATYILKTTCTCDAFAARLAGANRY 863  
QY 743 DOLKQGVGLSRVSAL-----PRDYGLEVGTWLNKLTGLGGAMRYEGKIRATAE 793  
Db 864 QRTENTPNCPTPGSFMGYTNTQPPRLATNLTAGLFFDQALTLGGRTY-TSQPTATAD 922  
QY 794 ERYIDGTNGGNTSNFRQLGRKRSIKQETTLARQPLIFDYAAEYKKNLIFRAEYKNLFR 853  
Db 923 KPWQVG-----ATTPOIEYRSVQ-----LPDFELKYLFEHTELNASLQNLTD 966  
QY 854 RYIDPLDAGNDAATQRYYSSEDFPKDKEDVTCNADKTLGNGKYGKGTGSKSVLTNFRGRTF 913  
Db 967 YYLDPL-----AQSFMP-----APGRTL 984  
QY 914 LITMSYKF 921  
Db 985 RVGMOAKF 992

RESULT 13  
US-09-897-516-4927  
; Sequence 4927, Application US/09897516  
; GENERAL INFORMATION:  
; APPLICANT: Corbin, David R.  
; APPLICANT: Goldman, Barry S.  
; APPLICANT: Hinkle, Gregory J.  
; APPLICANT: Huesing, Joseph E.  
; APPLICANT: Krasomil-Osterfeld, Karina C.  
; APPLICANT: Malvar, Thomas M.  
; APPLICANT: Slater, Steven C.  
; APPLICANT: Spiridonov, Sergei  
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof  
; FILE REFERENCE: 38-21(51847)B  
; CURRENT APPLICATION NUMBER: US/09/897,516  
; CURRENT FILING DATE: 2001-06-29  
; PRIOR APPLICATION NUMBER: US 60/215, 161  
; PRIOR FILING DATE: 2000-06-30  
; NUMBER OF SEQ ID NOS: 8409  
; SEQ ID NO 4927  
; LENGTH: 364  
; TYPE: PRT  
; ORGANISM: Xenorhabdus sp.  
US-09-897-516-4927

Query Match 8.4%; Score 410.5; DB 22; Length 364;  
Best Local Similarity 29.5%; Pred. No. 8.1e-29;  
Matches 113; Conservative 54; Mismatches 125; Indels 91; Gaps 12;

QY 74 ENLDNIVRSIPGAFQODKSSGIVSLNIRGDSGFRVNTWVDGITOTFYSTSDTAGRAG 133  
Db 39 ESMDSVLRSLRUPGYTQMDTSQGTIANVIRMGSGFRVNMVMDVGSQSFYIAPSEFAHGT 98  
QY 134 S--SQFASVDSNFIAGLDVVKGFSAGINSAGSANLRTLGLVDVVOGNNTYGLLLK 191  
Db 99 QPNYOTGALIDSFIIRTDVDRGOANDSDSNALVGSANFTIGDIDVIEGNKLGILTK 158  
QY 192 GLGTGNTSKNAMAAGAR-KWLESASGVLYGHSSRRSVAQNYRVGGGQHIGNFGAEY 250  
Db 159 SAYGTNGLKNGMIATAGRTQAFSTEGSIGAMLAISGHSIDAHYKNAM----- 207  
QY 251 LERRKQRYFVQEGALKFNDSGKWERDLQKQWKYKPKYKNYNNQELQKYLEGHDKSHREN 310

```
Db 208 -----VSSEEGT-----KTFNREPNSQLMKI-----N 231
QY 311 LAPQYDITPIDPSSLKQOSAGNLFKLEVDGVE--NKYTAQFRDLNTKIGSRKIINRYQF 368
Db 232 IKP-----NDFHELSGRFYHNKT-----KRHIDSVDYIL 263
QY 369 NYGLSLNPYTNL---NLTAAYNSGRQKYPKSGKFTGWLKLLKDFETYNNAKILDNLNTATF 425
Db 264 KY--HYTPFSELIDTNILLGSGKNQYFVKMSGLGKG-----ESHNKSNTIDIKNTSRF 316
QY 426 RLPRETELQTLGFNYFHNK 448
Db 317 NY-GETDISFTLGLSKLMDTEYHK 338

RESULT 14
US-60-215-161-4927
; Sequence 4927, Application US/60215161
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)A
; CURRENT APPLICATION NUMBER: US/60/215,161
; CURRENT FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 4927
; LENGTH: 364
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-60-215-161-4927

Query Match 8.4%; Score 410.5; DB 26; Length 364;
Best Local Similarity 29.5%; Pred. No. 8,1e-29;
Matches 113; Conservative 54; Mismatches 125; Indels 91; Gaps 12;

QY 74 ENLDNIVRSIPGATQODKSSGIVSLNIRGDSGFRVNTMVDGITQTFYSTSTDAGRAGG 133
Db 39 ESMDSVLSPLPGTYTQMDTSQGTIAVNIIRGMSGFRVNMVDGVSQSFYGIAPSEFAHG 98
QY 134 S--SFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVVOGNNNTYGLLLK 131
Db 99 QPNYQTGALIDSFIITDVRGOANDSDSVNALVGSANFRTIGIDDVIFEGNKLGLTK 158
QY 192 GLTGTNSTKGNMAAIGAR-KWLESGASVGVLVGHSSRRSVAQNYRVGGGQGHGNGAEY 250
Db 159 SAYGTNGLKNGMIATAGRTQAFSTEGSICAMLAISGHSIDAHKNAMG----- 207
QY 251 LERRKORYFVOEGALFNSDGKWERDLQROQWKYPKYNNOELQKYIEGHDKSWREN 310
Db 208 -----VSSEEGT-----KTFNREPNSQLMKI-----N 231
QY 311 LAPQYDITPIDPSSLKQOSAGNLFKLEVDGVE--NKYTAQFRDLNTKIGSRKIINRYQF 368
Db 232 IKP-----NDFHELSGRFYHNKT-----KRHIDSVDYIL 263
QY 369 NYGLSLNPYTNL---NLTAAYNSGRQKYPKSGKFTGWLKLLKDFETYNNAKILDNLNTATF 425
Db 264 KY--HYTPFSELIDTNILLGSGKNQYFVKMSGLGKG-----ESHNKSNTIDIKNTSRF 316
QY 426 RLPRETELQTLGFNYFHNK 448
Db 317 NY-GETDISFTLGLSKLMDTEYHK 338

RESULT 15
```

```
US-09-897-516-6415
; Sequence 6415, Application US/09897516
; GENERAL INFORMATION:
; APPLICANT: Corbin, David R.
; APPLICANT: Goldman, Barry S.
; APPLICANT: Hinkle, Gregory J.
; APPLICANT: Huesing, Joseph E.
; APPLICANT: Krasomil-Osterfeld, Karina C.
; APPLICANT: Malvar, Thomas M.
; APPLICANT: Slater, Steven C.
; APPLICANT: Spiridonov, Sergei
; APPLICANT: Hinkle, Gregory J.
; TITLE OF INVENTION: Xenorhabdus sp. Genome Sequences And Uses Thereof
; FILE REFERENCE: 38-21(51847)B
; CURRENT APPLICATION NUMBER: US/09/897,516
; CURRENT FILING DATE: 2001-06-29
; PRIOR APPLICATION NUMBER: US 60/215, 161
; PRIOR FILING DATE: 2000-06-30
; NUMBER OF SEQ ID NOS: 8409
; SEQ ID NO 6415
; LENGTH: 937
; TYPE: PRT
; ORGANISM: Xenorhabdus sp.
US-09-897-516-6415

Query Match 7.7%; Score 376.5; DB 22; Length 937;
Best Local Similarity 19.5%; Pred. No. 6e-25;
Matches 212; Conservative 150; Mismatches 380; Indels 343; Gaps 40;

QY 27 DAGRAGSEAOI-OVLEDVHVHAKRVKDKKVFYTDARAVSTRQDIFKSSSEMLNIVRSIPG 85
Db 6 DAGEGYDAVYDKDISNIYIGKKEIER-----YKGASPAD-VIRGAVG 47
QY 86 AFTQQDKSSGIVSLNIRGDSGFRVNTMVDGITQTFYSTSTDAGRAGGSQFASVDSNF 145
Db 48 VYSGDARNSGALDINIRGQGGRIPIVTIDGTEQ-----AITVGRYNGANNRYIDPNL 102
QY 146 IAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVVOGNNNTYGLLLKGLTGTNSTK----- 200
Db 103 ISSIEIEKGPSLNRVRKVGSGGAVSIKTLNIDDVPEGETFGINTKLETSSNSVKERTPS 162
QY 201 -----GNAMAAIGARKWLESGASVGVLVGHSSRRSVAQNYRVG-CGGQHIGN 245
Db 163 LSLGQDYRDVPNIHNGIESDPALKITPHSSKDNKLFKFDNA---FRVAVGTROEYFD 218
QY 246 FGAEYLERKQRYFVOEGA-----LKEN----- 268
Db 219 LMLAYAYRHKNYFAGKGGAHRYDAAATEADINLMLNPKYSLDPLPFAARIYRPGNEVP 278
QY 269 SDGSKWERDLQROQWKYPKYNNOELQKYIEGHDKSW-----RENLAPOY 315
Db 279 NTSSKMQSVLKNTHFT-----DEQALQLAPNTRMEFGDIMPSRLASVLAKENSVPQW 333
QY 316 DITPIDPSSLKQOSAGNLFKLEVDGVFNKY-----TAQFRDLNTKIG-SRKLIINRN 365
Db 334 PL-----ANARQQAASLNLYK--WDSANPYIDFDMNLWTTRTISNTSGGYPRGPVTRD 386
QY 366 YQF--NYGLSLNPYTNLNTAAVNSGRQKYPKSGKFTGWG--LLKDFETYNNAKILDNLN 421
Db 387 YDWEGGKGRSTIDGTLINTAVTNAQNNR-----WGVDISNKFELTQNLDLTLMGN 437
QY 422 TATFRLPRETEL-----OTTLGFNY----- 441
Db 438 FQERLGSDDDSHIDNLYFFQSPARKGQEQEINLAFNEDWRPTSWALGAGAKRYSYWS 497
QY 442 ----FHNE--YGNRPFPEELGLFF-----DGPD----QDNGLYSYIGR----- 474
Db 498 KODFLNERRIARDRNYKEPGEIIGKMSYWRVTVEDEANDFKRNDKEYYSKLSRTEKRE 557
QY 475 -----FKGDKGLLPQKSTIVO-----PAGSOYF-----NTFFYDAAL 506
Db 558 LRAKTKKIKDRNDRIQOKRWLVEEKFEWKYNPETGKLNKSDNPYFNQGLDMNEKVIDPIS 617
```

```
Qy 507 KDIYRLNYSNTVGYRFGGEYTGYYGSDDEKFRAGFGENSPTYKKKHONRSCGIYEPVLKK 566
Db 618 GREAYKEYSANSL-----SNDKILUSQDVDP-----WEPAPKR 652
Qy 567 YGKRANHSVSI SADFGDYFMPFASYSRTHRMENIOE--MYFSOIGDSGVHTALKPERA 624
Db 653 --KSHAWAPTFSATAYITDDLRYVTRYAEAVRMPISIFEDIVGFSGVKENYIGLKFPERA 710
Qy 625 NTWQFGFNTYKGLLKQDDTLGLKLVYRSRIDNYI--HNYYGKWMDLNGDIPSWVSSTG 682
Db 711 KTIETGFYDFSQLVNAERNADIKLSYNTVIENVFDRDNTY----- 752
Qy 683 LAYTIQHRNEK--DKVHKHGFELNVDYGRFFTNLSYAYOKSTQPTNFSAS--ESPNN 738
Db 753 -----NFAQLDKQKLAGLELQARYDNGSFFTDMLVYNMKNKYCDNNSAARMDSQNR 804
Qy 739 ASKEDQLKQY--GLSRVSALPRDYGRLEVGTWRGLGNKLTIGGAMRYFGKXSIRATAERY 796
Db 805 YGVPECIDGGFPGGYLTSIQPKYTANLVGGRLFDEKLELGSRLYHSRA--ENKDEKW 862
Qy 797 IDGTNGGNTSNFROLGKRISIKQETTLARQPLIFDFYAAEYEPKKNLIFRAEVKNLFDRIYI 856
Db 863 LMGV---LPNEYKGISNNPMRWNSVFT-----VDAYVSYQITPAISMELTGTNLNRYYL 914
Qy 857 DPLDAGNDAATQRYSSFDPKDKDEDVTCNADKTLNCKYGGTSGKSVLTNPFARGTFLIT 916
Db 915 DPL-----TRSMIP--APGRTFKLS 932
Qy 917 MSYKF 921
Db 933 LTSQF 937
```

Search completed: July 24, 2002, 08:58:34  
Job time: 450 sec







Db 213 IGGAEALLIERTGRAGEIRAHEAAGR-----GVQSFNRLAPVDDGSKYAYFIVEE- 263  
QY 284 KYRKYNNNOEL-----QKVEIHKD-----SWRENLAQVDITPIDPSSLKQOS-----AG 331  
Db 264 -----CKNGHEKCKANPKDVGEDKROTVSTRDYTGPNRFLA--DPLSVESRWLFPRG 317  
QY 332 NLF--KLEYDGVNKNYTAQ---FRDLNKTIGSRKII-----NNRYQF-----N 369  
Db 318 FREENKRYHIGGILERTQOTFDRMTVPAPFLTKAVEDANQKQAGSLRGNGNHYAGNHK 377  
QY 370 YGSLNPNYTNLNTAAYNSGRQYPKSGFTGWLKLDKFTYNNAKLLDNLNTATRLPR 429  
Db 378 YG-----GLTSGENNAAPGAAY-GTGVFYD----- 402  
QY 430 ETELOTTLGNFYHNEVGKRNFPPEELGFPD--GPQDNGL----- 468  
Db 403 ETHTKSRYLEYVYNADKDTWADYARLSYDRQDGLDNHFFQTHCSADGSKYCRPSAD 462  
QY 469 --YSYLGRRGDKGLLPQKSTIVQAGSYFNTFYFDAALKKDIYRLNYSNTVGY-REG 525  
Db 463 KPFSY---YKSDRVYIGESHKLLQAAPKKSFTAKI-----RHNLNVN-LGYDRFG 509  
QY 526 G-----EYTGYSDDDFKRAFGENSP----- 547  
Db 510 SNLRHODY--YYOSAN---RAYSLKTPQNNKKTSPNGREKNPYWVSGRGNVVTROIC 564  
QY 548 -----TYKKHCNRSC--GIYEPVKK-----YGKKRANHHSVS----- 578  
Db 565 LFGNNYTDCTPRSINGKSYAAVRDNVRLGRWADVAGLRYDYRSTHSDGGSVSTGTHR 624  
QY 579 -ISADFGDYMFPFASYSRTH-----RMPNIOEMFYSQIGDSGVHTALKPERANTWQFG- 630  
Db 625 TLSNAGILVKPADLWLDLTYRTSGRFLPSFAEMYGWRS GDKIKAVKIDPEKSFNKEAGI 684  
QY 631 -----FNTYKGLLKQDDTLGLKLVGYSRDN-----YTH----- 661  
Db 685 VFKGDFGNLEASFENAYRDLIVR-----GYEAQINDGREQVKGPNPAYLNAQSARI 735  
QY 662 ---NVYGGK--W---WDLNGDIP--SWYSSTGLAYT-IQHRNFKDKVHKHGFELNLDYGR 711  
Db 736 TGINILGKIDWGVWD---KLPGWST--FAYNRVRVDIKRADRTDIQSHL-FD--- 786  
QY 712 FFTNLSYAYOKSTOPTNFSDASEPNNAKEDQLKOGYGLSRVSALPRDYGRLEVGTWRL 771  
Db 787 -----AIOPSRYV-----VGSYDQPEGKGVNGMLT-----YSKAKEITELL 824  
QY 772 GNKLITLGMARYGKSRATABERYIDGTNGNTSNFRQLGKRSIKOTELARQPLIFDF 831  
Db 825 GSRALLNG-----NSRNTKATARRTRPWIYDV 852  
QY 832 YAAVEPKKLIIFRAEVKNLFRRYI 856  
Db 853 SGYYTVKKHFTLRAGVYNLNRHYV 877

RESULT 2  
US-09-584-501A-12  
; Sequence 12, Application US/09584501A  
; GENERAL INFORMATION:  
; APPLICANT: Lo, Reggie Y.C.  
; APPLICANT: Schryvers, Anthony B.  
; APPLICANT: Potter, Andrew A.  
; TITLE OF INVENTION: PASTEURELLA BINDING PROTEINS OF  
; PASTEURELLA HAEMOLYTICA AND VACCINES CONTAINING THE SAME  
; FILE REFERENCE: A34763 021645.0106  
; CURRENT APPLICATION NUMBER: US/09/584, 501A  
; PRIOR FILING DATE: 1996-11-29  
; PRIOR FILING DATE: 1996-11-29  
; PRIOR FILING DATE: 1996-11-29  
; PRIOR FILING DATE: 1995-12-01  
; PRIOR FILING DATE: 1995-12-01

; NUMBER OF SEQ ID NOS: 68  
; SOFTWARE: FastSeq for Windows Version 4.0  
; SEQ ID NO 12  
; LENGTH: 908  
; TYPE: PRT  
; ORGANISM: Neisseria meningitidis  
US-09-584-501A-12

Query Match 5.5%; Score 269; DB 5; Length 908;  
Best Local Similarity 21.2%; Pred. No. 2,3e-12;  
Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;

QY 5 FRLKPICFYLMGVTLHYSYAEDAGRAGSEAOIQVLEDVHVRAK--RVPKDKKVFTDARA 62  
Db 7 FRLNLCLSLMTALPV--YAENV--QAEQAQEKOLDTIQVRAKKQKTRREDNEVTGLGKL 61  
QY 63 VSTRQDIFYSSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFRVNTVVDGIT 118  
Db 62 VKSSDTLSKEQVLNIRDLTRYDPGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGVS 117  
QY 119 Q-TFYSTSTDAG--RAGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLG 175  
Db 118 QIOSYTAQALGTRTAGSSGAINIEYENKAVEISKGSNSSEYNGALASVAFQTKT 177  
QY 176 VDDVVOGNNYTGILLK-GLTGTNSTKGNMAAIGARKWLESASGVLYXGHSRRSVAQNY 234  
Db 178 AADIIEGQWGIQSTAYSGKDHALTQSIALAGR---SGGAELLIIYTKRRGREIHAAH 233  
QY 235 RVGGGQHQHGNFAEYLERKQ-----RYFVOEGAL-----KFNSD-SGKWER-- 276  
Db 234 KDAGKG--VQSFNRLVDEDKKEGSGYRYFIVEECHNGYAAACKNKLKEDASVKDERKT 291  
QY 277 -----DLQROWKYKP-----YKNYNNOELQKYIEGHDKSWRENLAPO 314  
Db 292 VSTQDVTGSNRLANPLEYGSQSWLFRPGWHLNDRHYVGAVLERTQOTEDT--RDMTVPA 349  
QY 315 YDITPID--PSSLK-----QOSAGNLFKLEY-DGVFNRYTAQFRDLNFK 355  
Db 350 Y-FTSEDYVPGSLKGLGKYSKDNKAERLFOGSGSTLGGIGYGTGVF-----YDERHTK 402  
QY 356 IGRKTIINRYOPNYGLSNPYTNLNLTAAYNSGRQYKPGSKFTGWLKDKDEFTYNNAK 415  
Db 403 -----NR-----YGVY-VYHNADKDTWADYARLSYDRQ----- 431  
QY 416 ILDLNNTATFLPRETELOTTLGFNFYHNEYGKNRPEELGLFFDGDQDNGLYSYLGRF 475  
Db 432 -IDLN-----RLQOT---HCSHDSKNCRP-----DG-----NKPYSF---Y 463  
QY 476 KGDGKLLPQKSTIVQAGSOYFNT-----FYFDAAL----- 506  
Db 464 KSDRMIEESRNLFOAVFKAFDTAKIRHNLINLGYDRFKSQLSHSDYVYLNQAVQYDL 523  
QY 507 -----KKDIYRLNYSNTVVG---YRFGGE-YT-----GYGSDDEPK 539  
Db 524 ITPKKPPFPNGSKDNPYRVSIGKTTVNTSPICRFGNNTYDCTPRNIGNGYAAVQDNV 583  
QY 540 RAFGENSPYKHKCNESCGIYEPVLKKYKKRANHSVS-----ISADFGDYFMPFASY 593  
Db 584 R-----LGRWADVAGIRYDYRSTHSDKSVSTGTHRNLSWAGVYVLPKFTWM 631  
QY 594 SRTH-----RMPNIOEMFYSQIGDSGVHTALKPERANTWQFG-----FNT 633  
Db 632 DLTYRASTGFRLPSPFAEMYCWARGESLKTLDLKPEKSFNREAGIVFKDGFNLEASYFNN 691  
QY 634 YKGLLKQDDTLGLKLVGYRSRDN-----YIH-----NVYGGK--WDLNG 672  
Db 692 AYRDLI-----AFGYETRTQNGQTSASGDPGVYRNAQNAIAGINILGKIDWGVWG 742  
QY 673 DIPSWVSSSTGLAY-----TIQHRNFKDKVHKHGFELNLDY--GREFTNL 716  
Db 743 GUPDGLYST-LAYNRKIKVKDADIRADRTFVTSYLFDAVOPSRVYVGLGIDHDPGIGINT 801

539	Db	KLKVGDIYVRDFDGVMSGVFAQ-----GEYNRDKLS	571
577	Qy	VSIADFGDYFMPFASYSRTHRPNIQEMFYSOIGDSGVHTALKPERANTWQF--GFNTYK	635
572	Db	AFISGSYSN---TGYWRYDR-----FYDK-----AHA-----KSKTVNIGWNA--	608
636	Qy	KGLLKQDDTTLGLKL---VGRSRIDNVIHNVYKWKWDLNGDIPSWSVSSTGLATYIQRNF	692
609	Db	KGGLNYNLTEHNHNFNANIGYISR-----APFFSGCAPLNSTVSNATN	650
693	Qy	KDKVHKHGFLELNLDYGRFFTNLSY--AYO-----KSTOPTNPSDASES--PNNASK	741
651	Db	PDAVNEKVFSEIGYGRSSEFLVINAYHTRMWDKTTTTRSQDITNYEGSLSEPDASK	710
742	Qy	EDQLKQGYGLSRVSALPR---DY-----GRLEVGT--RWLGKNKITLGGAMRYFG	785
711	Db	LVSTKSVINMGVNALLHQGVLEDFVAKPFQWLDSLGMFSGIGNRWDSN---ASGSFTVEG	767
786	Qy	KSIRATAERYIDGTNGCNTS-----	806
768	Db	QFVNSAS-----IKGSDGKDVTLVNAANAANGLEPQTMKLNKLKDVKGSGSAQTTAALGATEK	823
807	Qy	-----NFROLGKRKSIKOTETTLARQPL--IFFDFAVAAYE---PK	838
824	Db	IDKALRGIDWNLYARNYADWSLNSNDLVNSEKDFSTPHRIPTASTFDLNASYKKNFGK	863
839	Qy	KNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSF	874
884	Db	INAVISGNVNI.PQTVISDAPDGSNNHDKWTAYNVF	919

```

RESULT      4
US-09-545-199F-30
; Sequence 30, Application US/09545199F
; GENERAL INFORMATION:
; APPLICANT: Lowery E., David
; APPLICANT: Fuller E., Troy
; APPLICANT: Kennedy J., Michael
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions
; FILE REFERENCE: 28341/6227.NCP
; CURRENT APPLICATION NUMBER: US/09/545,199F
; PRIOR FILING DATE: 2000-04-06
; PRIOR APPLICATION NUMBER: 60/153,453
; PRIOR FILING DATE: 1999-09-10
; PRIOR APPLICATION NUMBER: 60/128,689
; PRIOR FILING DATE: 1999-04-09
; NUMBER OF SEQ ID NOS: 165
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 30
; LENGTH: 564
; TYPE: PRT
; ORGANISM: Pasteurella multocida
US-09-545-199F-30

Query Match          4.2%; Score 204.5; DB 5; Length 564;
Best Local Similarity 20.0%; Pred.No.1.3e-07;
Matches 131; Conservative 87; Mismatches 271; Indels 165; Gaps

Qy    41 EDVHVAKRVPKDKKVFTDARAVSTRODIKFSSLENLDNIVRSIPGAFQTQQDKSGGI VSLN   100
       ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| 
Db    44 EDNLPGSRVLUKNQID-----QQADNAADLINILPEVNAGGFPPGGGTLN 91

Qy    101 IRGD SGGRVTNWMDGITQTITFYSTDAGRAGGSQFGASVDNSFIAGLDDVKGSFSGA 160
        | | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| 
Db    92 INGMGDAEDVRQLDGATKSPEK-----YQGSIFFIEPELLRKRVTDKGNYSPQY 141

Qy    161 GINSLAGSANRLTGLGVDDVVQGNNTYGILLKLGTWS----TKGMAMAIGARKWLES 215
        | | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| 
Db    142 GNNGFAGTVKEETKDADTFDLKENOKI GLLFK--YGNNSNNQKTYSTATLVONEKNIDL 199

Qy    216 GASVGVL YGHRRSRAQNRY-----VGGGGQHIG-----NFGAEYLERRKKORYFVQ 261
        | | :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| 
```

Db 200 -----LLFGSVRN--ASNYTRPDKSKILFSKNNOKSLIKVNNQITPEHLLTLSSVYGIH 252  
Qy 262 EGALKFNSDSKWERDLQROQWKYKPKYKNYNQOELQYIEGHDKSWRENLAPOYDITPID 321  
Db 253 KG-----WEPWAARVMSRP-----TETBIKH--GIDVAKRKL-- 287  
Qy 322 PSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRKLIINRYQFNGLSLNPNYTNLN 381  
Db 288 -----YRD-----QKDESYSLKRYLPPENKNWINLS 313  
Qy 382 LTAAYNSGRQYKPGSKFTG--WGLL--KOFETYNNAKILDLNNTATFRPRETELOTTL 437  
Db 314 VOLSYTEQNDTRHEKVTFSFLTGLNKSWITYSDL--TFDISNTSLNIGR--AEHELLF 371  
Qy 438 GFNFHNEYGNKRRPEELGLFFDDPDODNGLYSYLGRFKGDKGLLPQKSTIVOPAGSOYF 497  
Db 372 GLOWLKNK--RNTLMYHKG-----GVKKADYNYGFQY-----YMPSGRQYT 412  
Qy 498 NTFYDAAKDKDIYRLNYSTNTVGYRFGGEYTYGYSDDDEFKRAFGEN--SPTYKKHCNRS 556  
Db 413 QAFVLODOIKNONFL-----TGGIRY-----DHINNIGOKNLAPRYN--DIS 453  
Qy 557 CGIYEPVLKYGKGRANNHS--VSIADFGDYFMPFASYSRTHRMPNIQEMY----FSQIG 611  
Db 454 AG-----HDYSQNYNGWSYILGKYDVNHYLSLFTNFSKTRAPVIDEQYETQYSQAS 507  
Qy 612 DSGVHTALKPRANTWOFENTYKKGKLLKODDGLGLKLVGRSRIYHNHYG 665  
Db 508 VSATSLNEREMINQTRVGIIITLNLHFOENDAFQFTTYFYNRCKNEIFKTRG 561

RESULT 5  
US-09-540-209B-9521  
; Sequence 9521, Application US/09540209B  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES FR  
; FILE REFERENCE: 2709.1001-001  
; CURRENT APPLICATION NUMBER: US/09/540, 209B  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 10444  
; SEQ ID NO 9521  
; LENGTH: 1083  
; TYPE: PR  
; ORGANISM: B.fragilis  
US-09-540-209B-9521

Query Match 4.1%; Score 200.5; DB 5; Length 1083;  
Best Local Similarity 20.2%; Pred. No. 6.8e-07;  
Matches 197; Conservative 129; Mismatches 369; Indels 279; Gaps 57;  
Qy 4 SFLRKPICFVLMGVTLHYHYAEDAGRAGSEAOIQLV--EDVHVAKRV---PKDKKV- 56  
Db 86 SLKVAPNSQFLVS-----YIGKQOTIKVSESTYNIIVKEDAEVLDEVVVGVSQKKVN 141  
Qy 57 -----FTDARAVSTRQDIFKSSSENLDNIVRSIPGAFTQODKS-----SGIVSLNIRGDSG 106  
Db 142 VTGAVGHVSAEALR--PVANASQALQGV--PGLNLTVCNNGGALDGLTNMIRGACT 197  
Qy 107 FGRVN-----TWDGIGTQTYSTSD-----AGRAGGSQFGASVDSNFIAGLDVVKG 154  
Db 198 IGDGSGSPLVIDIGIEGLDNTNPNIDIESVSLKDAASASIYGARASFGVI--LVTTKS 255  
Qy 155 SFGSAGINSIAGSANLR--TLGVDDVVOGNNTYCLLLKGLTGINTKGN-----AMAA 206  
Db 256 GKSKNTV--SYSGSARFSAIGVPM--DSY--TFAQYFNRAANKGGDIFAPAYMER 310  
Qy 207 IGA-----RKWLESASVG---VLYGHSRRSVAQNYRVGGGQHIGN 245  
Db 311 IKAQEGTLKATTVYDNGAGIWKWANANGTDWFEFYYDHWASQEHNLNSINGTD----- 366

Qy 246 FGAELYERRKORYFV-----QEGALKFNSD-----SGK-----WERDLQROQWKY 285  
Db 367 -----KTQYLISGSFLDQKGLMRHGDKFQRYTLNGKIITAVTDWFKVYISTKWR 417  
Qy 286 KPKYNNQOELQYIEG---HDKSWRENLAPOYDIT--PIDPSSLKQOSAGNLFKLEYDG 340  
Db 418 EDF-----ERPSTLGTGNFFHNLARKVPVHPAYDPNGFPMDGEVEQEMGGKQNSQKDF 471  
Qy 341 VFNKYTAQFRDL--NTKIGSRKLIINRYQFNGLSLNPNYTNLNTAAYNSGRQYKPGSKF 399  
Db 472 YTNOLQVLFPPIKNNKINLDSVRTTQYOHWEVLVY-----AYNAVADPY-----Y 519  
Qy 400 TGGLGLKDEFTYNNAKILDILNNTATFRPRETELOTTLGFNFHNEYGNKRRPEELGLFF 459  
Db 520 TVMDMGVGYAAGSSRVN-----YSWKENYTTTNIYSDFYKQF- 558  
Qy 460 DGPDDQNGLY--SYLGRFKGD---KGLLPQKSTIVOPAGSOYFNTF-----YDAA 505  
Db 559 -----DNGHYFKVMAGFNABLYKTRNITAEKNTLITP--GVPTINTATDDPQAYGGYADNS 612  
Qy 506 LKDDIYRLNYSTNTVGYRFGGEYTYGYSDDDEF--KRAFGENSPYTKKHCNRSQGIYEPV 563  
Db 613 VAGFFARVNNWSYKD---RYMFEANGRYDGSRVGVKRWG--FPFSFAGWNIA---REPF 665  
Qy 564 LKYGKGRANNHNSVISADFG-----DYFMPFASYSRTHRMPNIQEMYFSQIGDSGV 615  
Db 666 MESFAEK--INMGLKLRAWSGQLGNTNTDAWYF-----YQTMP--VGSNYGWLNVGERP 718  
Qy 616 HTALKP-----ERANTWQFG-----FNTYKKGKLLKODDGLGLKLVGRSRIYHN 661  
Db 719 NYATNFGIVSSKKTWETVETWVGLWDSFNNRLSG-----SFDFYFVR 761  
Qy 662 NVYKGWDLNGDIPSWVSSYGLAYTIOHRNFKDKVHKHKGFELELNY-----DYGRFFT 714  
Db 762 YTY----DMIGPAPELSSLLGT--SVPKINNDS--MKSYGFELEVNWRDRDICEVSYGAKFV 814  
Qy 715 NLSYAYOKSQPTNFSADSSPNNAKEDOLKOGYGLSRV-----SALPRDYGRLE 765  
Db 815 --LSDQOKILR---YPNDSYDVGSYKGEHLNDIWLTTTIGIAKSOEEMDAHLAKVDSS 870  
Qy 766 VGTFRW--LG-----NKLTLGAMRVFGKSIKIRATAERY---IDGT--NGG 803  
Db 871 VGTNMGVGDIMYADLDGDGKISNGTNKLGTDGYRIIGNS---TPRFKYGITLDAWKGF 927  
Qy 804 NTSNFRQ--LGKRSI 816  
Db 928 DFSIFMQGIGKRD 941

RESULT 6  
US-09-540-209B-6249  
; Sequence 6249, Application US/09540209B  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES  
; FILE REFERENCE: 2709.1001-001  
; CURRENT APPLICATION NUMBER: US/09/540, 209B  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 10444  
; SEQ ID NO 6249  
; LENGTH: 925  
; TYPE: PR  
; ORGANISM: B.fragilis  
US-09-540-209B-6249

Query Match 4.1%; Score 200; DB 5; Length 925;  
Best Local Similarity 19.8%; Pred. No. 5.9e-07;  
Matches 154; Conservative 107; Mismatches 257; Indels 260; Gaps 40;  
Qy 34 EAQIOVLDEHVHAKRVKPKDKKVFYTDARAVSTRQ-----DIFKSSSENLDNIVRS 82

Db 236 EETHMLDEVITISGRIONVKSTOLGNETIRPTQLKNIPMALGEVDILK-----WVQA 288  
Qy 83 IPGAFTQODKSSGIVSINIRGDSGFRVNTWVGITOTFYSTS-----T 126  
Db 289 LPGVKTYGEASSG---ENVRGGATDQNLILLNDG---TIYNPNHLFGFFAAFNDSMVKEA 342  
Qy 127 DAGRAGSSGOFASVDSNF--IAGLDVVVKGSFSGSAGINSLAGSANLRTLGVDG---VVQ 181  
Db 343 EYKSSIPAQGGRISSILDTKEANKERTGSAGIGLVTSKLNLEIPIIKORTSVLLS 402  
Qy 182 GNNYTG-LLKGLGTGNTSK-GNA---MAAIGARKWLES-GASVGLVYGHSSRRSVAQNY 234  
Db 403 GRTTYSIDWIMKQPEKSEYKNGTAGFYDLAAIVAHKFNDSNLVNYGYSHDRFAFNSNE 462  
Qy 235 RVGGGGHIGNFGAEYLERKQRYFVQEGALKFNDSGKWERDLQROQWKYKPYKNYNNQ 294  
Db 463 KYG-----YNNLNA-----SARWR-----AVENEK 482  
Qy 295 ELOKYEIGHDKSWRENLAPOYDITPIDPSLKOOSAGNLFKLEYDGVFNKYTAQFRDLNT 354  
Db 483 LIGYFSAGYD-----HYD-----NNRETVNASTAKLSFD--INQTF-----V 519  
Qy 355 KIGSRKIINRYOFNYGLSNPYTNLNLTAAYNSGROKYPKSGKFTGWLKDKDFETYNNA 414  
Db 520 KADFTNILADKHTLNFGEK-----SMLYHINSGYE-PEGS-----ESFVK 560  
Qy 415 KILDNLNT--ATFRLPRETELQTLGNYFNHNEYGKNRFPPEELGLFFDGDQDNGLYSYL 472  
Db 561 DVLOKDALETAFLVGDWEITPKLSVN-----AGIRY-----SLFSAL 599  
Qy 473 G---RFGKDGKLLPOKSTI--VOPAGSOYENTFYDAAALKKDIYRLNYSNTVGYRFGG 526  
Db 600 GPRSYYOYAGMLPHESTIITDTITAGAGKFMKTYH-----GP 636  
Qy 527 EY--TGYGSDDEPKAFGENS-----PTYK-----550  
Db 637 EFLRSARYAFTDNFVSKAGNSMRQYIHKLSNTVIMSPDTWKLSDVNIKPRGQWQAAG 696  
Qy 551 KHCNRSGIPEVLKYKGRKRRNNHVSISADFGDYPMPFASRTRHRMNIQPMWFSQI 610  
Db 697 LYLNSPISGWEYSVEGYKRR-----MSDLYD-RGGAKLIMNHIET--DVINTQG 744  
Qy 611 GDSVHTALAPE--RANTWQFGFNTYKGLLKODDITGLKLVGRSRIIDNYIHNVYKWW 668  
Db 745 HAYGVELOVKQKVGKLGMM--SVTYSRTFLQND-----KRIEKPVNN--GDWY 790  
Qy 669 DLNGDIPSWSSSTGLAYTIQHRNFKDKVHKHGFLELNYDYGFRFT-NLSVATQKSTQ 725  
Db 791 PTEYDKPHDFKFGV-----NVK-FTHRYSMSINVDYSTGRPTTIPAGQYVDESTQ 839

RESULT 7  
US-09-545-199F-32  
; Sequence 32, Application US/09545199F  
; GENERAL INFORMATION:  
; APPLICANT: Lowery E., David  
; APPLICANT: Kennedy J., Troy  
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
; FILE REFERENCE: 28341/6227.NCP  
; CURRENT APPLICATION NUMBER: US/09/545,199F  
; CURRENT FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: 60/153,453  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: 60/128,689  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 165  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 32  
; LENGTH: 967  
; TYPE: PRN  
; ORGANISM: Pasteurella multocida  
US-09-545-199F-32

Query Match 4.0%; Score 194.5; DB 5; Length 967;  
Best Local Similarity 19.0%; Pred. No. 1.7e-06;  
Matches 207; Conservative 131; Mismatches 339; Indels 413; Gaps 51;

Qy 39 VLEDVHVHAKRVKDKKVFDDARAVSTVRODIFKSSE-NLDNIVRSIPCAFTQODKSSGIV 97  
Db 22 ILADSHOEATEL-----DTITVSSQDDEMTKEKIGETVKTASOLKQKQVQDSDRL 73  
Qy 98 SLNIRGDSGFRVNTWVGITOTFYSTSDTAGRAGSSGQFASVDSNFIA-----GLDVV- 152  
Db 74 --VRYETGV-----TVVEAGRFGSGYAIRGVNDENRVAITVDGLHQA 114  
Qy 153 -----KGSFSGSAGINSLAGSANLRTLGVDVVOGNNTYGLLLKGLGTGNTSKG---- 201  
Db 115 TSSQGFKELFEGYGNFNTRNSVEIETLVAKIAKADSVK-----GSGSLGAVLF 168  
Qy 202 --NAMAAGARKW--LESGASGVLYGHSSRVAQNYRVGGGQHGIGFAGYLERK 255  
Db 169 ETKDARDFLTEKDMHIGYKAGYSTADNOGLNAVTLAGYQM-----FDALIMHSKR 219  
Qy 256 QRYFVQEGALKFNDSGKWERDLQ--ROQWKYKPYKNYNNQLOKY-----299  
Db 220 HGHELE---NYDYKNG---RDIQKEREKADPYTITKESTLVKFSFSPTEHHRFTVASD 272  
Qy 300 -----IEGHDKSW-----RENLA---POYDITPI-----320  
Db 273 TYLQHSRGHDLSYNLVATHTHQLDEKESRHANDLTKRKNVSFTYENTVTFPMDTKLSY 332  
Qy 321 -----DPSSLKQOS-----AGNLFKLEYDGVFNKYTAQF 349  
Db 333 SQORITTRATEDYCDNEICDSYKNPLGLQFDGQILDPAKNKIKOGSGL-----STQI 388  
Qy 350 RDNLTKIGSRKIIINRYQFNYSLSL-----NPTNMLNLTAAYNSGRQ---391  
Db 389 VDENGKFPPTTGTGNNAAFSNNLRLRTGFWLDCVFDCKNPFVYVINSNGTYOAREVLL 448  
Qy 392 -----KYPKSGKFTGWL-----LKDFETYNNAKILDLNNTATFRL 427  
Db 449 SEETVQDGLYKTAKEG-GLPNYLILPNNSKGYLPYDYKERDLNNTKQINLDTKTF-- 505  
Qy 428 PRETELTTLGNYFHN-EYG--KNREPEEL--GLFEDG-----PDODNGLYSYLG--R 474  
Db 506 -----LTFNLENLSYGGVYSRIEKEINKAGYGRNPTWADRILGQSSYCGYNA 556  
Qy 475 FKGDKG-----LILPQKSTIVQAGSOYFNTFYDAAKKDIYRLNYSNTVGYR----- 523  
Db 557 LKCPKHEPLTSFLIPVEATT-----QSLYFANILKVH-----NMISIDLGYRYDHIK 603  
Qy 524 FGGEYT-----GYI-----GSDDEFKRA-FGENSPYKHKHNRSCGIYE 561  
Db 604 YNPEYTPGVTPKIPDDMVKGLFIPMPKEPOLKDFDYAKFGEAYKKWKEYLPKNA--EE 661  
Qy 562 PVLKYYGKRRANNHVSISADFG--DYMPFASYSRTRHRMNIQEMYFS-QIGDSGV--H 616  
Db 662 NIAYIAQDKTKFKHYSLSGATFDPNLNRLVQVYKSKFRAPTSDELYFTFKHPDFTILPN 721  
Qy 617 TALKPERANTWQFGFNTYKGLLKODDITGLKLVGRSRIIDNYIHNVYKWWDLNGDIPS 676  
Db 722 PVLKPEEAKNOETALTVDHNNWGFVSTSVFQTK---YRHFID-----LA 761  
Qy 677 WVSSTGLAYTI---QHRNFK-----DKVHKHGFLELNYDYGFRF-----TNLSY 718  
Db 762 YLGSRLNSNSVGGQAQARDQVYQNVVNDNAKVGLKLEINARNLGLYFHWHLGDPNTSYKF 821  
Qy 719 AYQKSTOPTNFSASESPNNA-----SKEDQLKQGYGLSRVS-----A 756  
Db 822 TYQGRILD-----GDRPMNAIQPKASVFGLYGDYDHKENKFGADLYITRVSEKAKADPNM 875  
Qy 757 LPRDYGRLEVGTRWLGKNTLGGAMRYFGKSIRATAERYIDGTNGTNGTNSFRQLGKRSI 816  
Db 876 FYKEQGYKDSAVRWRSDDYTLVDVAVGI-----903

QY 817 KOTETLARQPLIFDYAAYEPKKNLIFRAEYKYNLFDRRY-----IDPLDAGN----- 863  
Db 904 -----RPIKNLTLOFGVYNLTDRKYLTWESARSIKPFGTSLINQK 944  
QY 864 -DAATQRYYS 872  
Db 945 TGAGINRFYS 954  
RESULT 8  
US-09-540-209B-9746  
; Sequence 9746, Application US/09540209B  
; GENERAL INFORMATION:  
; APPLICANT: Gary L. Breton  
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR  
; FILE REFERENCE: 2709.1001-001  
; CURRENT APPLICATION NUMBER: US/09/540, 209B  
; CURRENT FILING DATE: 2000-04-04  
; NUMBER OF SEQ ID NOS: 10444  
; SEQ ID NO 9746  
; LENGTH: 783  
; TYPE: PRT  
; ORGANISM: B.fragilis  
US-09-540-209B-9746

Query Match 3.7%; Score 179.5; DB 5; Length 783;  
Best Local Similarity 20.8%; Pred. No. 1.9e-05;  
Matches 139; Conservative 84; Mismatches 251; Indels 193; Gaps 30;  
QY 34 EAQIQVLEDVHVAKRVPDKKVTDAVSTRTQDIFPKSSNLDNIVSRIPGAFTOODKS 93  
Db 107 EPTQSLGSEVVYAKSEARQLREQAPMPSVISMQLOGTVSNVDVLSKTVGVTIRTTGG 166  
QY 94 SGIVS-LNIRGDSGFRVNTWVGITOTFTYSTSTDAGRAGSSQFGASVDSNFIAGLDVV 152  
Db 167 VGSSRSRVSRGLEG-KRIGFFIDGSPMNDNSDFIDN-----DIPVDMIDRIEY 215  
QY 153 KGSFSGSAGINSLAGSANL-----RTLGVDVVOGNNYTGILLKGLTGTNSTKGNAMA 205  
Db 216 KGVVPAREGSSVGGAVNIYVIREPPKYLDASYSIESFNTHKLSL-----VTKRNI-- 266  
QY 206 AIGARKWLESASGVLYGHSRRSVAQNYRVGGGQHGNGFAEYLERRRQRYFVQ--E 262  
Db 267 --ATKGLEFGG--GGFYTS--DNNYKME-----SPFEEGLIIRKRNHDKFKKLAVA 311  
QY 263 GALKFNSDSKWERDLORQWKKYP-----YK-----NYNNQELQKY----- 299  
Db 312 GSLK---ARKWFDLA---EFEPVFIHTFKETQIEYNIERAKHTYSDAFIFANKLEKE 363  
QY 300 ---IEGHDKSWRENLAPOYDI-TPIDPSSLKQOSAGNLFKL--EYDGVFNKYTAQFDLN 353  
Db 364 NFLTEGLD--WESNLAYATVFHMVDTAARYNWDGTYTPAVSEYGEIGKWNASARN-- 419  
QY 354 TKIGSRKIINRNQYGLSLNPTNLNTAAYNSGRQKYPKSKFTG----- 402  
Db 420 ----EKHTIHLNVIINNHSINLSLFSASGHPKDDLANKVVGKYTNFSTWASW 475  
QY 403 --GLLKDFETYNNAKILDLN-----NT-----ATFRLPRETELOTTLGFN 441  
Db 476 IAGLGYDFRTDNDIFLSNLAVKYMYGMNTHMSSINSSEAEKVDMLKRDGFSNALRYR 535  
QY 442 FHNEYGK-----NRPEELGLEFDGPDQDNGLYSYLGRFGKGLLPKQSTIVQAGSQ 495  
Db 536 TPDPMGLSVGYDVRLPAESSELGDC-----YTVAPSGN--LLPERNTSV----- 578  
QY 496 YFNTFYDAALKKDIYRLNYSNTVGYRFGGVTGYGSDDEFKRAFGENSPYKKHCNR 555  
Db 579 --NLGFLLDRTKXASNLQVEVNT-----FYGYLENMIRFTG----- 613  
QY 556 SCGIYEPVLKKGKRRANNHVSISADFGYFMPFASYS-----RTH----- 597

Db 614 --CYLOSQYQNFCKMRTLGVEVEVKADLTHWLYGYNMTYQDLRDVRKKEPNTHITNPTK 671  
QY 598 --RMPNI 602  
Db 672 GSRMPNI 678  
RESULT 9  
US-09-545-199F-105  
; Sequence 105, Application US/09545199F  
; GENERAL INFORMATION:  
; APPLICANT: Lowery E., David  
; APPLICANT: Fuller E., Troy  
; APPLICANT: Kennedy J., Michael  
; TITLE OF INVENTION: Anti-Bacterial Vaccine Compositions  
; FILE REFERENCE: 28341/6227.NCP  
; CURRENT APPLICATION NUMBER: US/09/545,199F  
; CURRENT FILING DATE: 2000-04-06  
; PRIOR APPLICATION NUMBER: 60/153,453  
; PRIOR FILING DATE: 1999-09-10  
; PRIOR APPLICATION NUMBER: 60/128,689  
; PRIOR FILING DATE: 1999-04-09  
; NUMBER OF SEQ ID NOS: 165  
; SOFTWARE: PatentIn Ver. 2.0  
; SEQ ID NO 105  
; LENGTH: 669  
; TYPE: PRT  
; ORGANISM: Pasteurella multocida  
US-09-545-199F-105  
Query Match 3.6%; Score 177; DB 5; Length 669;  
Best Local Similarity 20.3%; Pred. No. 2.3e-05;  
Matches 170; Conservative 95; Mismatches 296; Indels 276; Gaps 40;  
QY 109 RVMTWVDGTTQTFYSTDA-----GRAGSSQFGASVDSNFIAGLDVYVKGFSFGSAGINS 164  
Db 8 RVAVIVDGIPIQAEISTISARYSTERHNGNIN--NIEVENVSSLKVQKGAASVMYGSGA 64  
QY 165 LAGSANLRTLGVDVVOGNNYTGILLKGLTGTNSTKGNAMAIG-----AR 210  
Db 65 LGGTVEFTTKDIEDFVEPGHHLGFLSK--FGYTSKNREYRQVIGVGKGGEHFGFVOLT 122  
QY 211 KW----LESASGVLYGHSRRSVAQNY-----RVGGGQHGNGFAEYLERRRQRYF 259  
Db 123 RWGHETINNGKGDILGEHRGKPNLNYTTSMITKVGYDINNTHRF-TLFLDREKKL 181  
QY 260 VQEGALKFNSDSKWERDLORQWKKYPKYNYNNOELQYIEGHDKSWREN-----LAPQY 315  
Db 182 TEEKTLGL-SDAVRFAND-----QTPYLRYGIE--YRY--NGLSWLETVKLFLAKQ- 227  
QY 316 DITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFDRLNTKIGSRKIINRNQYFNYGLSLN 375  
Db 228 -----KIEORSALQEF-----DIN-----NRN----- 244  
QY 376 PVTNLNLTAAY-NSGRQKYPKSKFTG---WG-----LLKDFETYNNAKILDLNNTATFR 426  
Db 245 ---KLDSTMSFVYLOQNTARGESTSPLYWGPSRHLSAKFE-FRDKFLENKHKHFTFR 300  
QY 427 LPRETELOTTLGPNYFNHNEYKGNRPPELGLFFDGPQDQDNGLYSYLGRFGKGLLPQKS 486  
Db 301 -PMQ-----INRFRQO--GRNNYTE-----LGRWDHY----- 317  
QY 487 TIYQVAGSQYFNTFYFDAALKKDIYRLNYSNTVGYRFGGVTGYGSDDEFKRAFGENS 546  
Db 318 --VFPVKRSFEFSLSLMDDDTKIGELLHLG-----LGRWDHY----- 351  
QY 547 PTYKHCNRCSCGIYEPVLKKGKRRANNHVSISADFGYFMPFASY--SRTHRMPIIOE 604  
Db 352 -NYKPLNLSQHNINRTORLPY-PKTSKFSYQLSLEYQLHPSHOIAYRLSTGERVRVED 409  
QY 605 MYFSQIGDSG-----VHTALKPERANTWQFGFNTYKGLLKKODDTGLGLKLVGYRS 654

```
Db 410 LYFEDRCKSSQFLPNPDLOPEALNHEISYRFQNOYAHESVGL-----FRT 456
QY 655 RIDNYIHNVYKGMWDLNGDIPSWVSTGLAYTIOHRNEK-DKVHKHGFELNAYDYGR-- 711
Db 457 RYHNF-----IQEREMTCDKI-----PYEYNRTYCYCT 484
QY 712 -----FETNLSYAVOKSTOPTNFSDAESPNNASKEDQLKQGYCLSRVSALPRDYGRLEV 766
Db 485 HNTYVMEVNEFAVIGKEVSGALNGSAGFLSDGLTFRLKGSYSKGO-----NHDGDLKS 540
QY 767 GTRWL-----GNKLTGLGAMRY-FGKSTRATAERY IDGTNGGNTSNFRQLGKRS 815
Db 541 IQPTVVVTGIDYETEGHSVLSG--RYSAAKAKADAETEYTH-----DKKV 585
QY 816 IKOTETLAROPLIFDYAAVEPKKNLIFRAEVKNLFDRIYDPLDAGNDAATQRYYS 872
Db 586 VKQPHLSPSYFVVDFTGQVNLNKGVLNMGVENLFNRDYM-TWDSAYNLTRGYTS 641

RESULT 10
US-09-540-209B-10058
; Sequence 10058, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FR
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540, 209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 10058
; LENGTH: 707
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-10058
```

```
Query Match 3.6%; Score 176; DB 5; Length 707;
Best Local Similarity 20.7%; Pred. No. 3e-05;
Matches 174; Conservative 94; Mismatches 333; Indels 238; Gaps 47;
```

```
QY 36 QIOVLEDVHVYKAKRVKDKKVFVDARAVSTRQDIFKSSNLDNIVRSIPGAFQDQKSSG 95
Db 1 EVQI-----VSTRATSTPVAFTNVSKELKQNF--GQDIPFLSMTPSALATSDAG 53
QY 96 I--VSLNIRGDSGFRVNTWVDGITQT-----FYSTSTDAGRAGGSGOFASVDSNEI 146
Db 54 IGYTLVRGTG-D-TRINITANGIPMDAESHTLFWNMPD-----PASSVKD--- 100
QY 147 AGLDVVKSGSPSGAGINSLAGSANLRTLGVDDVVQGNNTYGLLLKGLTGTNSTKGNAMAA 206
Db 101 --IQVORGACTSTNGAGAFGASVNMQTEGI-----SNQPYAEINASVDSFNAHKETVKFG 153
QY 207 IGARK--WLESG--ASGVLYGHSSRV--AONTRVGGGGHIGNFGAEYLERKQRFTVQ 261
Db 154 TGLLKDHAFDARLSTGTGIDRASVDLYSEYAQGG-----YFAD 195
QY 262 EGALKNSDSGK-----WERDLQOQWKYK-----YKNYNQOELQ- 297
Db 196 NTSVKFTTFGGKEKYHAWYATKEEMKKYGPREFNSCCMTDHHGHRFYKQDQDNVLMQ 255
QY 298 --KYIEGH--DKSWRENLAPOYDITPID-----PSSLKQOSAGNLFKLEYDGVFNKY 345
Db 256 NYQLLNHTESAANLNAALHY--TKGDYGYQYKEDRSIKEY---RLHPFMYDG---K 306
QY 346 TAQPRDLNTKIGSKLIINRYQFNGLSLNPYTNLNLTAAYNSGRQXPKGSKFTGW--- 402
Db 307 EVEKSDL---IROKKMDN---HFGGGVFSVNYPHQNMDASLVNALNY-----DGHWF 354
QY 403 -----GLLKDEFTY--NNAKILD--LNNTATRLPRETELQTTLGFENFNE--YKG 448
Db 355 RVTWKVIYIGELLPDHEYYRNKAKKTDGNLYLKANYNLVAGLNAYADLQYRYNKHG- 413
```

## RESULT 11

```
US-09-540-209B-8695
; Sequence 8695, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540, 209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 8695
; LENGTH: 1094
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-8695
```

```
Query Match 3.5%; Score 173; DB 5; Length 1094;
Best Local Similarity 20.4%; Pred. No. 9.7e-05;
Matches 174; Conservative 114; Mismatches 288; Indels 276; Gaps 48;
```

```
QY 52 KDKKVFVDARAVSTRQDIFKSSNLDNIVRSIPGAFQDQKSSGIVSLNI--RGDSGFR 109
Db 100 KDKYIM-----IIPQSKVEVESKKLSGIVK-----DDKGPLIGVNVVSFKG-SPTGT 145
QY 110 VNTWVDG-----ITQTFYSTSTDAGRAGGSSQF-----GASVDSNFIAGLDV 151
Db 146 V-TGLDGRFSILAAGNIIEFSYGVYTYQITVGDASSLTVVLEEDAKALDEVVVVTALGI 204
QY 152 VKGSFSGSAGINSLAGSANLRTLGVDDVVQGNNTYGLLLK--GLTGTNSTKGNAMAAIGA 209
Db 205 KRAEKALSYSVQVKSDA-----INDVKDANFVNLTKGVAGVSINRSSGIG---GA 254
QY 210 RKWLESAS--VG---VLY-----GHSRRSVAQNYRVGGGGHIGNFGAEYLERK- 255
Db 255 TRVYMRGAKSIVGNNNNLYVVDGMPIGNPSKGEINNDSTPGGEGISDFNPEDIESLSI 314
QY 256 -----QRYFVQEGALAFNSDSGKWERDLQOQWKYKPYKNYNQOELQYI----- 300
Db 315 LTGPAAALYGVSSAANGVILINTKKG-----QEGKLKISINNTFETPTPYWPFQN 366
QY 301 -----EGHDKSWRENLAPOYDITPIDPSSLKQ-----QSAGNLFKLEYDGVFNKYTAQFRD 351
Db 367 RYGNKAGSYKSWGEMLQ-----QFSTFRPKDFFKTGANIMNAANFSVGNKNNQTFVS 418
QY 352 LNTKIGSRKII--NRNYQFNVGL-----SLNPYTNLNLTAAY-----NSGR----- 390
```

Db	419	VAT-TNSTGLIPNNEYRYNFTLRNTASMLNDKLHLDLGASYVLQGDQNWLSAGRYFNPL	477
Qy	391	--QKYPKSGKFTGWLKDDFTYNNNAKILDLNNTATFRLPRETELOTTLGFNYFHNEY-	446
Db	478	VPLYLFRGDEFA--VKYERY-----DTNR---KFPIQEWYSYQOGLN-LENPYW	523
Qy	447	-----CKNRPPELGLFFGPDQDNGLSYL--GRFGDKLLPKQ-----STIV	489
Db	524	IVNREMEVSKKRY-----MEYANVYD--ILSWLNIAGRVDNTTTSERKLHASTIK	576
Qy	490	QPAGS-----QVNTFYFDAAL--KKDIYRLVNSTN-----TVGYRFGGE--	527
Db	577	LHAQSKGAYNRSWEERYQOITYADIMLVNKNFGFNFLTANAGFSYEDHLLTGMGTGGKLF	636
Qy	528	-----YTYGYSDDDEKFRAPGENSPITYKKHCNRCGI-----YEPVLKKYKKKRANNHVS	577
Db	637	TVPNLFSAY-----NFDPASGPGCSHHTHRNNSVVFSTELCYKKSMLYLTLCROEWASQ	691
Qy	578	SISADFGDYEMPASYSRTHRMENIOEY-----FSOIGSGSVHTALKP--	621
Db	692	LVNSDQTYEYPSVGVGV-----ISEMWSLPKFISEWKNRASFSAEVGPGIINTGLTPGT	746
Qy	622	-----ERANTWQGFNTYKKGLLQDDDTGLKLVGYRSRIDN	658
Db	747	VTDPMKGGVINPISVYFPNPKAEQTSYELGTN-----LRLFNKNINIDA	792
Qy	659	--YIHNVYGGWMDLNDIPDSWVSSTGLA-YTHOHRNFKDKVHKHGFELNVD--YGR--	711
Db	793	TVLTDYTYNQT-----LSSMSPASGISGVYQ-----AGVRNKKGLESLIGTNRDFKGV	843
Qy	712	FFTNLSYAYOKS	723
Db	844	YATNLTYATANRN	855

## RESULT 12

```

US-09-540-209B-7143
; Sequence 7143, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTERIOIDES FRAGILIS
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540,209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 7143
; LENGTH: 833
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-7143

```

Query Match	3.5%	Score 172.5;	DB 5;	Length 833;
Best Local Similarity	19.8%;	Pred. No. 7.2e-05;		
Matches 187;	Conservative	120;	Mismatches 351;	
			Indels	285;
			Gaps	44;

Qy	210	RKWLESGASVGVLYGHGRRSVAQNYRVGGQGHGNFGAEBYELRRKQRYFVQEGALKFNS	269
		: :	
Db	269	YTHLKTGTV-----RANAFY	287
		: :	
Qy	270	DSGKWERDLQROQWKYPKYNNNOELQYIEGHDKSWENLAPOYDITPIDPSSLKQOS	329
		: :	
Db	288	DN-----DYKVFVPIIDLATNKI-----DERWVRFNDA-----RS	320
		: :	
Qy	330	AGNLFKLEYGVFNKYTAQFRDLNLTKIGRSKIIRNYQFNYGSLNPTYNLNLTAAYNSG	389
		: :	
Db	321	GG--IRLE--TGITNKPADY-----LLAGIILSKN-----DKDVQGTGATMDAVYGVG	364
		: :	
Qy	390	RQK-----YPKSGKFTGWLGLKDFNYNNAKILDLNNTA--TFRLPRELETOITLGF	439
		: :	
Db	365	KMKSESVIPSYIRYKKDOLFIDGSLUSLYGYNSVNTFVNDTITARRYNWLGESVPSISAGE	424
		: :	
Qy	440	NYFINEXGKNRFEELGFDPDQDNGLSYLGREFKGDKGLLPQKSTIVOPAGSQYFNT	499
		: :	
Db	425	GYYTDSKIKNR--EWLG-----NGNISYV-----IDGHQSILNLHVWSAMRRT	465
		: :	
Qy	500	FYFDAALKDIYRLNYSNTVGVYRFGEGEYGYGSDDEPKR-----AFGE-----NSPTYKK	551
		: :	
Db	466	M-----NDKVRPDDENNVPPOOLTKNITG--IGWQIRYDRWNANVFGKMYKLYSSTYKR	517
		: :	
Qy	552	HCNRSÇGYIEPVLLKY-----GKRRANNHSVSISADFGDYFMPFA-----SYSRTRHMPN	601
		: :	
Db	518	-----LDEYENARWEKVRDHKTNFGYGAAATVYILPSLQAKFSYEHAVALRPE	565
		: :	
Qy	602	IQEMYFSQIGDSGV-----HTALKPERANTWQFGFN-----TYKKGLLKODDTLGLKLGVYRS	655
		: :	
Db	566	SIEMF-----GDGLTQQRNPOLKPESSRNLLGLHSFTQTGAHQLSAD-----	608
		: :	
Qy	656	IDNVIHNVYKQWDLNGIDIPSWSSSTGLATYIOHRNFQKVKHKGHGFELNLNDY--GRFPT	714
		: :	
Db	609	-GNFIYR--YTTDFILKG-----VSLISNP--TTGYENL--GKVLTKGYEAARVRYNKDLFHT	659
		: :	
Qy	715	NLSVYQKSTQPTNFDASCS-----PNNASKEDOLKOGYGLSRVSPALPRDYGKLEVG	767
		: :	
Db	660	GAGETYQDITDRQRYEKTQDSFVGEGITENITYKE-----RLPNIPYLFANADAG	709
		: :	
Qy	768	TR-----WLGKMLTGGAMDYF-----GKSIATAEBRYI--DCTNGGNTSNFRQ	810
		: :	
Db	710	VRFHDLTWRNSVLTFDYNLNYIHUSYILSPFCLCAKSKKVIPOFQSHDLALGYSDN-----	766
		: :	
Qy	811	LGKRSIKOTETLARQPLIFDFAAYEPPKKNLIFRAEVKNLFR	853
		: :	
Db	767	-GKYSVVVECTNLTNOKLYDNYRLOKPGR--AFNVKLRYFFESK	806

RESULT 13

```

US-09-540-209B-8474
; Sequence 8474, Application US/09540209B
; GENERAL INFORMATION:
; APPLICANT: Gary L. Breton
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO BACTEROIDES
; TITLE OF INVENTION: FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: 2709.1001-001
; CURRENT APPLICATION NUMBER: US/09/540.209B
; CURRENT FILING DATE: 2000-04-04
; NUMBER OF SEQ ID NOS: 10444
; SEQ ID NO 8474
; LENGTH: 702
; TYPE: PRT
; ORGANISM: B.fragilis
US-09-540-209B-8474

```

Query Match 3.5%; Score 171.5; DB 5; Length 702;  
Best Local Similarity 19.68; Pred. No. 6.7e-05;  
Matches 149; Conservative 102; Mismatches 260; Indels 251; Gaps 39;  
Qv 170 NLRITGVDDVVQGNNTYGLLLKLGITGTNTSTKGNMAAIGARKWLESGASGVLYGHSRRS 229



```
Db 95 SLNVQVSDAVK--HFAGVTVKDYGGIGGLKTVLSLGAEH-----TAVG--YDGTIS 145
QY 230 VAQNYRVGGGQHIGNFGAERYL-----RKQYFVQEGALKFNSDSGKWERD 277
Db 146 DCQT-----GQIDIGFSLNDRLSLNSGOSDNIFQPARFFASAGIL-----N 189
QY 278 LQROQWKYKPYNNQELQKYTEGHDKSWRENLAPOYDITP---IDPSSLKOOSAGNLF 334
Db 190 IOTLTQFQKDN-----RTNLSAFKTSWGLVNPSSLQLEQKLSRW 231
QY 335 KLEVDG-----VENKYTAQFRLNTKIGSRKLIINRYQENYGL--SLNPTNLNT 383
Db 232 VLSANGEMWAGADHYPTLHYGEDNDLTSREKRNKTEVKNLRAEAGLFGNFSDEQWRK 291
QY 384 AAYNSGRQKYPKSGKSTGWLGLKDFETYNNAKILDNL-----NTATFRLPRETELQTLGF 439
Db 292 AYYOSSRGLPNATY-----YDYSOHLWDRNVFVOSQYKKEFSQWQFQSAKW 343
QY 440 NYFHNEYKGRNFPPEELGLFPDGDQDNGLYSYLGRFKGDKGLLPQKSTIVOPAGSQYFNT 499
Db 344 NWSYQRY-----LD-PD-----YKSGSEG-----KTNSYYQOE 370
QY 500 FYEDA-ALKKDIYRLNYSTNT-----VGYRFGGEVTVGYGSDDEFKRA 541
Db 371 YILSASALYRVLSNLSFSLDASINRLNANLKDFAPTRYSWLTAFAKYVND----- 424
QY 542 FGNSPTYKKHCRNSGCIYEPVLK---KYGKRRANHSVS---ISADFGDYFMPFAS----- 592
Db 425 -----WLTASASVLTATVINEEVRQSGSAAANRRKLSPPYSAS-----FKPFASEEPR 470
QY 593 -----YSRTHRMPIQIEMFYFSQIGDSGVHTALKPERANTWQFGNTYKGLLKQDDTLGLK 648
Db 471 IRLFYKDIFRLPSFNDLYQGVG-----NTNLKPESTTQYNLGL--TYSRST-----NE 517
QY 649 LVGYRS--RIDNYTHNVYKMWDLNGDTPS-----WVSSTGLAYTIQHRNFKDKVHKHGE 702
Db 518 LIPYVSTADAY-----YNKVKDIIAIPTKNLFIV--SWNLG-----KVDIKGID 562
QY 703 LELNYD---YGRFETNLS---YAYOKSTOPTNFSDESPPNASKEDOLKOGYGLSRVSAL 757
Db 563 IAGNISLQPEKRLVNLSGNVTYQAL-----DMTEPGGKTYKQ-----QIAYT 606
QY 758 PRDYGRLEVG--TRW--LGNKLLGLGAMRYFGKSIRATAEERIDTNGTNGTNSFRLQK 813
Db 607 PRVSGSQAGIETPPWNLSYSLFSGKRYMLGNLQNR-----ENRLDSYSDHSVSRSRLRI 662
QY 814 RSIKQETTLARQPLIFDFYAAEYKPKNLIFRAEYKLNLFDRRY 855
Db 663 RNVTSLTV-----EVLNLLDKNY 681
```

```
RESULT 14
US-09-584-501A-2
; Sequence 2, Application US/09584501A
; GENERAL INFORMATION:
; APPLICANT: Schryvers, Anthony B.
; APPLICANT: Potter, Andrew A.
; TITLE OF INVENTION: TRANSFERRIN BINDING PROTEINS OF
; TITLE OF INVENTION: PASTEURILLA HAEMOLYTICA AND VACCINES CONTAINING THE SAME
; FILE REFERENCE: A34763 021645.0106
; CURRENT APPLICATION NUMBER: US/09/584,501A
; CURRENT FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: 08/753,759
; PRIOR FILING DATE: 1996-11-29
; PRIOR APPLICATION NUMBER: CA 2,164,274
; PRIOR FILING DATE: 1995-12-01
; PRIOR APPLICATION NUMBER: 60/008,569
; PRIOR FILING DATE: 1995-12-01
; NUMBER OF SEQ ID NOS: 68
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 2
```

```
; LENGTH: 930
; TYPE: PRT
; ORGANISM: Pasteurella haemolytica
US-09-584-501A-2

Query Match 3.4%; Score 167; DB 5; Length 930;
Best Local Similarity 19.2%; Pred. No. 0.00033;
Matches 213; Conservative 142; Mismatches 375; Indels 380; Gaps 58;

QY 5 FRKPICFYLMGVTVLHYHYVAEDAGRAGSEAOIQVLEDV-----HVAKRVPKDKV 56
Db 8 FRYPVALTVLFALSHSYGAATENKKEENNDLAVLDEVIVTESHYAHERQNEVTGLGKV 67
QY 57 FTDARAVSTRQDIFKSENLDNIVRSIPG-AFTQQDK--SSGIVSLNIRGDSGFGRYNTM 113
Db 68 VKNYHEMSKQIQL-----GIRDLTRYDPGISVVEQGRGASSGYA---IRGVDK-NRYSLL 118
QY 114 VDGITO--TFYSTSTDRAGGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANL 171
Db 119 VDLGPOAHSYHTLGS DAN--GGAIN---EIEYENIRSIELSGASSAEYSGAHGGAIGF 173
QY 172 RTLGVDDVVOGNNTYGLLLK-GLTGTNSTKGNAMAAIARKWLESASGVLYVGHSSRSV 230
Db 174 RTKDAQDIIKEGQHWGLDSTSYASKNSHFLOSTAAAGEAGGFEALVIATHRHGKETI 233
QY 231 AQNYRVGGGQOHIGNFGAERYLERKQRYFVOEGALKFNSD-----SGKWERDLQROQW 283
Db 234 SEANKLKHNRITRTGTGFENRYDTQIPHRMLLEDLLIVEDTCPTLDCPTPRARVKNL 293
QY 284 KYPKYNNYNNQELQ-----KYIEGHDKSWRENLAPOYDITPIDPSSLKOOSAGNLF 334
Db 294 PVTFPEYTPPEERKQLEQIPYRTEQLSAQEYTGKDRIAP-----NPLDYKS----- 339
QY 335 KLEVDGVFNKYTAQF-----RDLNTKIGSRKLIINRYQENYGLSLNPTV---NL 380
Db 340 -----NSVFMKFGYHFNSSHYLGAILEDTKTRYDIRDQTPAY-----YTKDINL 385
QY 381 NLTAAYNSG-----ROKYPKSGKFTGWLKDFETYNNAKILDNNTATFRLPRET 431
Db 386 SLRNVYVGGNLDGLVFKPRIPYGLR-----YSHVKFFD-----ER 422
QY 432 ELQTLTGFNHYFNEYKGRNFPPEELGLFPDGDQDNGLYSYLGR----- 474
Db 423 HKRRRLGFTYKYP-ENNRWLDLSIKUSAD--KODIELYSRLHRLHCSDPVVDKNCRPTL 479
QY 475 -----FKGDKGLLPQKSTIVQ-----PAGSOYFN-TFYFDAALKKDIYR--LNTSTN 518
Db 480 DKSMSMYRTERNYQEKHRYVHLEFDFKALNAGOGVFNQTHKLNGLGDFRNSLMDHGM 539
QY 519 TVGYRFGGEYTVGYGSDDEFKRAFGENSPYKHH-----CNRSCGIY---EPVLKK 566
Db 540 TAQYTKGG-YTSYRG-----RGRLDNPIYRRDPRSIETVSLCNRTRGDLNCEP---- 588
QY 567 YGKRRANHSVSISADFGDYFMPFASYSRTHRMPIQIEMFYFSQIGDSGVHTALKPERANT 626
Db 589 -RKIKGDSHFVS-----FRLVISEYVDLGLGVRFQDHR--- 621
QY 627 WQFGENTYKKGLLKQDDTLGLKLVGYSRIDNYTHNVYKMWDLNGDI-----PSWVSSTG 682
Db 622 -----FKSDDPDWLTSRT-YRN-----WSWNGGTLTKPTEFVS--- 652
QY 683 LAYTIQHRNFKDKVHKHGFELNLYD-YGRFFTNLSYAYOKSTOPTNFSDESAPSNASK 741
Db 653 LSVRI-----SNGFRVPAPFELYGK-----RDHIGLKD-NEVYQRAQR 689
QY 742 EDQLKQGYGLSRVSALPRDYGRLEVGTRWGLNKLTLGAMRYFG--KSRATAEERYID 798
Db 690 SHOLE-----PEKSTNHEIGVSFKQFGYLD--VSIFRNKYNNKIATACKRIIQ 736
QY 799 GTNGGNTSNFRLQKRSIKQETTLARQPLIFD-----FYA--AYEPKKNLIFRA 845
Db 737 KSH--CFYNTHTQDVALNGINLVAK----FDLHGILSMLPDGIFYSSVAN-----RV 783
```





**THIS PAGE BLANK (18870)**

```
Db 128 VSFQERDDPATVINIRGLQDFGRVAVVVDGARQNYQT-----CHNANGSFF---LPEL 180
Qy 146 IAGLDVVKSGSSAGINSLAGSANTLRLLVDDVVQGNNTYGLLKLGLTGCTNSTKGNAMA 205
Db 181 IGGVDVVRGPTANTYGAIGGLVSFRKTDINDVLRPGERGVDSLGSYGNRRGLGSV 240
Qy 206 AIGARKWLESCASGVLYGHSSRSVAQNYRVGGGQHIGNGAEYLERRRKQRYFVQEGAL 265
Db 241 FGGVRAIPDVDFGAVY-----RTQNGYK-DGNGTEIGTNGQ-----VEAGLM 284
Qy 266 KFNSSDKWERDLQRQWKYKPKYKNYNQELQKYIEGHDKSWRENLAPOYDITPIDPSSL 325
Db 285 KLT-----VYP-----ALGHEV----- 296
Qy 326 KQOSAGNLFKLEYDGVFNKTAQFRDLNLTGSRKLIINRNQYFNGISLNPYTNLNTAA 385
Db 297 -----KFGAVFDQYQYDYG-----QFNRG-----PTTQAL-IA 324
Qy 386 YNSGRQKYPKGSK-FTG-----WGLLKD-----FETYNNAKILDLNNTATFRLPRETEL 433
Db 325 LNRSSVYASDAKNYSGFTWNYSLPSDLNLFDMHMSVYGNR--TDNDQTKTYHYGTTPSA 382
Qy 434 QTTLGF-----NYPHNEYG--KNRPPEELGLFFDGDQDNGLYSYLGRFKGD- 478
Db 383 YCNGGFGNNVSGVCDKRGYVLTNYGYDANNTR-----FNVGWRNALTWGVDAFQDDV 437
Qy 479 -----KGLLPQKSTIVOPAG-----SOYENTFYFDA--ALKKDIYRLNY-STNT 519
Db 438 ITTDSRG-----NSNITTPSGIRTVSGGFLQLKQNYST-WLEAVSAIRYDHYDLDSGKTST 492
Qy 520 VGYRFGYEITYGSDDEFKRAFGENSPYKHKHCRSCGIEPVLYKKGKRRANNHVSVI 579
Db 493 GDRFSPKIT-----LGV-TPV----- 508
Qy 580 SADFGDYFMPASYSRTHRPNIOEMVFSQIGDSGVHTA-----SGAATGGGPAFFVCPDGTAGLFCFLP 558
Db 509 -----PGFQPVVSAEGRASITETVI-----SGAATGGGPAFFVCPDGTAGLFCFLP 558
Qy 619 ---LKPERANTWQGFNTYKGLLKQDDTLGLKLGVRSRIDNYIHNYGKWMDLNGDIP 675
Db 559 NPNLRPEVGNKKEVINLKYDNIESANDSPRGKINLFRNDVSDVI-----DLVASAP 610
Qy 676 SWVSTG-LATYIGHRFKDKVHKHGELELNYDYGFFTNLSYAY-QKSTQPTNFSAS 733
Db 611 VAVPPGFSFQYQYQNIAN-ARIRGFEAETMYDAGDWFIVGAGHYIQGKNVATNIGLAT 669
Qy 734 ESPNASKEDQKQYGLSRVSALPRDYGRLEVGTWLGKNTLGGAMRYFGKSIRATAE 793
Db 670 ITP-----RKVVTGGVRLLDRLILITAQ 693
Qy 794 ERYIDGNGGNTSNRQLGKRSIKQETLARQPLIFDYAAEYKKNLIFRAEYKNLFD- 852
Db 694 WASYGPNDVPAGYLPATGYE-----LVNLYLTYNATRDIVLSASIDNLLNQ 740
Qy 853 ---RRYIDPLDAGNDAATQ 868
Db 741 YRYPYAIIP-GSSTDGTTQ 757

RESULT 15
Q9KLM6
ID O9KLM6 PRELIMINARY; PRT; 784 AA.
AC O9KLM6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DE 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DE 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TONB RECEPTOR-RELATED PROTEIN.
GN VCA0625.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
```

```
RP SEQUENCE FROM N.A.
RC STRAIN-EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,
RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tetelin H., Richardson D.,
RA Ermolaeva M.D., Vantheman J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Utterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae.";
RL Nature 406:477-483(2000).
DR EMBL; AE004392; AAF96526.1; .
DR TIGR; VCA0625; .
DR InterPro; IPR000531; TonB_boxC.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 784 AA; 87575 MW; 605DCFF12B0CBB10 CRC64;
```

Query Match 6.3%; Score 309.5; DB 16; Length 784;  
Best Local Similarity 21.2%; Pred. No. 5.2e-09;  
Matches 195; Conservative 122; Mismatches 343; Indels 261; Gaps 43;

```
Qy 6 RLKPICFYLMGVTLHYHYSYAEADAGRAGSEAQIOVLVDHVHAKRVPKDKVFTDARAVST 65
Db 36 QLKPL-FTLLPVLSSVVQAQE---NTEQAVDETVTVH-----QOSILTQDQ---T 79
Qy 66 RODIFKSSNDLNVRSIPGAFTQDKSSGIVS-----LNIRGDSGFGVRN 111
Db 80 RSDLDK-----VRGIANA---DIFSGITSVOSNMHNEAGALDIGIKGVQGEGRVP 127
Qy 112 TMVDGITQTFYSTDAGRAGSSQFGASVDSNFTAGLDVVVKG-----FSGSAGINS 164
Db 128 IFIDGSLQ---STHTRGYQGVSDR--TYIDTLLSSLTVNKGATIESSPYASGAVG--- 179
Qy 165 LAGSANLRTLGVDVVQGNNTYGLLLKG-----LTGTNSTKGN-AMAAIGARKWL 213
Db 180 --GVVNATTILGDIKDDQAFVVLKARANNNHNTPDVSGDYSEQOYALDERGHSF 237
Qy 214 ESGA-----SVGVLYGHSRVSVAQNYR-----VGGGQGHIGNFGAEYLE 252
Db 238 KHGSLMLGLGVAQASFTNLVAYSRRKSNHFGAGKGYEYQEPVVGQGEVNTSFE--- 294
Qy 253 RRKORYFVQEGALKFNSSDGKWERDLQRQWKYKPKYKNYNQELQKYIEGHDKSWRENLA 312
Db 295 -----SDSWLFKIASDTG---TAHNADFNRYHQAQKAGEVLWAYWYKSSDEWGNPY 343
Qy 313 POYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYYTAQFRDLNLTGSRKLIINRNQYFNGL 372
Db 344 PD-----GKDRMPQWGLGT-----AKVNTYSA-----NYYPQPDH-- 373
Qy 373 SLNPYTNLNLTAAYNSGRQKYPKGSKFTG-WGLLKDPETYNNAKILDLNNTATFRLPRET 431
Db 374 ---PWLNLNANFWYTEADL-----AQYNGLWALGTNABOYFAY---HNRSGSLTNET 422
Qy 432 EL-QTTLGFNYFHNEYGKNRPPEELGLFFDGDQDNGLYSYLGRFKGDKLLPKOKSTIVQ 490
Db 423 LLTQMPVRLNGLAQNERLSPEE-----DGQT-----RE-----TKTVTS 458
Qy 491 PAGSOYFNFTFYDAALKDIYRLNYSNTNYGVYRFGGEYTYGSDDEPKRAFGENSPYK 550
Db 459 RHGKRTAQNLAFANADIDYSPLRVQLGLNLHNAK-----STDYQTKQOL-----DYK 504
Qy 551 KHCNRSCGIEPVLYKKGKRRANNHVSISADFGDYFMP-----FASYSTRTHRPNIOEMY 606
Db 505 E-----KDLDSLEFVALTPSPQLFLKLSRTYRMPSLYETT 540
Qy 607 FS-QIGDSGVHTALKPERANTWQGFNTYKGLLKQDDTLGLKLGVRSRIDNYIHNYG 665
Db 541 LSNEFVSYPNPNPKPEQAWNNEVGQVPMASNSVLQDRLNLSYFNSLKDFISG--G 598
```





```
PS Claim 3; Page 91-95; 103pp; English.
XX The present sequence represents a BASB024 outer membrane protein of
CC Neisseria meningitidis. The BASB024 polynucleotide sequence was
CC first identified in the Incyte Pathoseq database containing
CC unfinished genomic DNA sequence of N. meningitidis. BASB024
CC polypeptides and polynucleotides are useful for generating an
CC immune response in an animal. Antibodies specific BASB024 polypeptides
CC are useful for treating N. meningitidis infection, which causes
CC bacteremia and meningitis.
XX Sequence 921 AA;
SQ
Query Match 100.0%; Score 4894; DB 21; Length 921;
Best Local Similarity 100.0%; Pred. No. 0;
Matches 921; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MRSSFRLKPCFYLMGVTLHYHYSAEDAGRAGSEAOIQVLEDVHVAKRVPKDKKVFETDA 60
DB 1 mrssfrikpicfymgvtylhysyaedagragseaqlqvledvhvakrvpkdkkvvftda 60
QY 61 RAVSTRODIFKSSENLDNIVRSIPGAFQODKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
DB 61 ravstrdqifkssenldnivrsipgafqodkssgivslnirgdsgrvntwvdmvgitqt 120
QY 121 FYSTSDAGRAGSSQFASVDSNFIAGLDVVKGFSFGSAGINSLAGSANLRTLGVDVV 180
DB 121 fyststdagragssqfagsvdsnfiaqldvvkgsfsgsaginslagsanltrlgvddv 180
QY 181 QGNNTYGLLLKGLTGTNSTGNMAAIGARKWLESGASVGLYGHSSRVAQNYRVGGGG 240
DB 181 qgnntyglllkgltgtnstgnmaaiarkwlesgasvglghssrsvaqrnyrvgggg 240
QY 241 QHIGNFAEYLERKQRYFQEGALKFNSDSGKWERDLQOQWKYKPYKNVNNQELQYI 300
DB 241 qhignfaeylerkrqryfvegealkfnsdsgkwerdlrqoqwkypkyknynnqelqyi 300
QY 301 EGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAOFRDLNKTIGSRK 360
DB 301 eghdkswrenlapdyditpidpsslkqgsagnlfkleydgvfnkyytaqfrdlntkigsrk 360
QY 361 IINRNYQFNGLSNPTNLNLTAAVNSGRQKYPKSGKFTGWGLLKDFEYNNNAKIIDLNL 420
DB 361 iinrnyqfnvqlsnptnlnltaaavnsgrqkypksgkftgwglldkfetynnakiidl 420
QY 421 NTATFRLPRETELQTLGCFYFHNKYGNRPFPEELGLFFDGPQDNGLYSVLGRFKDGK 480
DB 421 ntatfrlpretelqtltgcfyfhnekygnrpfpeelglffdgpdnglysvlgrfkdgk 480
QY 481 LLPQKSTIVOPAGSQYFNTFYFDAALKKDIYRLNYSNTVYGRFGGEYTYGYSDDDEFKR 540
DB 481 llpqkstivopagsqyftntfyfdaalkkdiyrlnysntvygrfggeytygyssdddefkr 540
QY 541 AFGNSPTYKHKCNRSQGIPEVLKYGKGRANNHVSISADFGDYFMPFASYSRTHRMP 600
DB 541 afgnsptykhhcnrscgipevlkylgkgrannhsvsisadfgdyfmpfasystrhrmp 600
QY 601 NIQEMVFSQIDGSGVHTALPERANTWQFGFNTYKGLLKDDPTLGLKLYGRSRIDNYI 660
DB 601 niqemyfsqidsgvhtalperantwqfgfntykgllddptlglklygrsridnyi 660
QY 661 HNVYCKWMDLNGDIPSWSVSTGLAYTTQHRNFKDKVHKHGFELNLDYDGRFFTNLSYAY 720
DB 661 hnvycwmdlndgipswsvstglayttqhrnfkdvkhkgfelnldydygrfftnlsyay 720
QY 721 OKSTQPTNFSDASESPNNAKEDOLKOGYGLSRVSALPRDYGRLEVTWGLNKLTLGGA 780
DB 721 okstqptnfdsasespnnaakedolkogyglsrvsalprdygrlevtwglwnkltlga 780
QY 781 MRYFGKSI RATAERYIDGTNGGNTSNFROLGKRISIKQETLAROPLIFFDYAAYEPKN 840
DB 781 mryfgksirataeeryidgtnggntsnfrqlgkrsikqetlarqpliffdyayepkn 840
```

```
QY 841 LIFRAEVKNLFDYRIDPLDAGNDAAATQRYSSFPDKDKEDVTCNADKTLGNGKYGTS 900
DB 841 lifraevknlfdrdyridpldagndaatqryssfdpdkdkedvtcnadkltlcnkygts 900
QY 901 KSVLTNPFARGRTFLITMSYKF 921
DB 901 ksvltnpfargrtflitmsykf 921
RESULT 2
AAY69381
ID AAY69381 standard; Protein; 922 AA.
XX
AC AAY69381;
XX
DT 19-JUN-2000 (first entry)
XX
DE A BASB024 outer membrane protein of N. meningitidis.
XX
KW BASB024; outer membrane protein; N. meningitidis infection;
KW bacteremia; meningitis.
XX
OS Neisseria meningitidis.
XX
PN WO200011182-A1.
XX
PD 02-MAR-2000.
XX
PF 13-AUG-1999; 99WO-EP05989.
XX
PR 18-AUG-1998; 98GB-0018004.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
WP I; 2000-224702/19.
DR N-PSDB; AA261581.
XX
PT Novel polypeptides derived from the products of the BASB024 gene of
PT Neisseria meningitidis, useful for inducing an immune response and
PT producing antibodies useful for treating meningitis -
XX
PS Claim 3; Page 88-91; 103pp; English.
XX
CC The present sequence represents a BASB024 outer membrane protein of
CC Neisseria meningitidis. The BASB024 polynucleotide sequence was
CC first identified in the Incyte Pathoseq database containing
CC unfinished genomic DNA sequence of N. meningitidis. BASB024
CC polypeptides and polynucleotides are useful for generating an
CC immune response in an animal. Antibodies specific BASB024 polypeptides
CC are useful for treating N. meningitidis infection, which causes
CC bacteremia and meningitis.
XX
SQ Sequence 922 AA;
Query Match 97.7%; Score 4779.5; DB 21; Length 922;
Best Local Similarity 97.9%; Pred. No. 0;
Matches 903; Conservative 7; Mismatches 11; Indels 1; Gaps 1;
QY 1 MRSSFRLKPCFYLMGVTLHYHYSAEDAGRAGSEAOIQVLEDVHVAKRVPKDKKVFETDA 60
DB 1 mrssfrikpicfymgvtylhysyaedagragseaqlqvledvhvakrvpkdkkvvftda 60
QY 61 RAVSTRODIFKSSENLDNIVRSIPGAFQODKSSGIYSLNIRGDSGFRVNTWVDGITQT 120
DB 61 ravstrdqifkssenldnivrsipgafqodkssgivslnirgdsgrvntwvdmvgitqt 120
QY 121 FYSTSDAGRAGSSQFASVDSNFIAGLDVVKGFSFGSAGINSLAGSANLRTLGVDVV 180
DB 121 fyststdagragssqfagsvdsnfiaqldvvkgsfsgsaginslagsanltrlgvddv 180
```



```
QY 181 QGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESASGVLYGHSRRSVAQNIRVGGG 240
Db 181 qgnntyglllkgltgtnstkgnamaaigarkwlesasgvlyghsrrsvaqnirvsgg 240
QY 241 QHIGNFGAEYLERRKORYFVQEGALKFNSDSGKWERDLQROQWKYPKYNKN-QELQKY 299
Db 241 qhignfgaeylerrrkoryfvqegalkfnsnsgkwerdfqrpywktyqkynndpqelqky 300
QY 300 IEGHDKSWRENLAPOYDITPIDPSLSKQOSAGNLFKLEYDGVFNKYTAQFRDLNFKIGSR 359
Db 301 ieghdkswrenlapqyditpidpslskqosagnlfkleydgvfnkytaqfrdlntkigsr 360
QY 360 KIINRNQFNYGLSLNPTNLNLTAAYNSGRQKYPKSGKFTGWLGLLKDFEYNNAKIIDL 419
Db 361 kiinrnyqfnyglslnsyanlnltaaynsgrqkypksgkftgwglldkfetynnakildl 420
QY 420 NNTATFRLPRETELQTLTGPNYFNEYCKNRPPEELGIFFDGPDQDNGLYSLGRFGDK 479
Db 421 nntatfrlpretelqtltgfnymeygnrffpeelgiffdgpddnglyslgrfgdk 480
QY 480 GLLPKQSTIVOPAGSQYFNTFYDAALKKDIYRLNLTNTVGYRFGGEYTYGYGSDDEFK 539
Db 481 glpkqstivopagsqyftntfydaalkkdiyrlnystntvgyrfggeytygygsddefk 540
QY 540 RAFGENSPYKKHCNRSRGYIPEVVLKYGKRRANNHVSISADFGDYPMPFASYSRTHRM 599
Db 541 rafgensptykhhcnqscgiyepvvlkyygkrrannhsvsisadfgdyfmpfasyrthrm 600
QY 600 PNIQEMTFSQIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 659
Db 601 pniqemyfsqigdsghvtalkperantwqgfntykkglkqddtlglklvgyrsridny 660
QY 660 IHNYGKWDJUNGDIPIPSVSVSTGLAYTIQHRNFKDKVHKHGFLELANYDYGRFFTNLSYA 719
Db 661 ihnygkwdjungsipsvsvstglaytiqhrnfkdvkhkhgflelnydygrfftnlsya 720
QY 720 YOKSTQPNFSDASEPNNAKEDQLQGYGLSRVSALPRDYGRLVGTWHLGNKLLILGG 779
Db 721 ykstqpnfnsdasepnnsakdqkqgyglsrvsalprdygrlevgtwrlgnklltgg 780
QY 780 AMRYFGKSIRATAEERYIDGPNNGNTSNFRQLGKRSIKQTETLARQPLIFDFYAAEPKK 839
Db 781 amryfgksirateaeryidgntngntsnvrqlgkrsikqtetlarqplifdfyaaepkk 840
QY 840 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLCKNGKYGT 899
Db 841 nlifraevknlfdriryidpldagndaatqryysfdpkdkdedvtcnadkctlcngkygt 900
QY 900 SKSVLTNPFARGTRTFLTMSYKF 921
Db 901 sksvltnfargtrtfltmsykf 922

RESULT 3
AAV69380
ID AAV69380 standard; Protein; 922 AA.
XX
AC AAV69380;
XX
DT 19-JUN-2000 (first entry)
XX
DE A BASB024 outer membrane protein of N. meningitidis.
KW BASB024; outer membrane protein; N. meningitidis infection;
KW bacteremia; meningitis.
OS Neisseria meningitidis.
XX
PN WO200011182-A1.
XX
PD 02-MAR-2000.
XX
```

```
PF 13-AUG-1999; 99WO-EP05989.
XX
PR 18-AUG-1998; 98GB-0018004.
XX
PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.
XX
PI Thonnard J;
XX
DR WPI; 2000-224702/19.
DR N-PSDB; AAZ61580.
XX
PT Novel polypeptides derived from the products of the BASB024 gene of
PT Neisseria meningitidis, useful for inducing an immune response and
PT producing antibodies useful for treating meningitis -
XX
PS Claim 5; Page 83-86; 103pp; English.
XX
CC The present sequence represents a BASB024 outer membrane protein of
CC Neisseria meningitidis. The BASB024 polynucleotide sequence was
CC first identified in the Incyte Pathoseq database containing
CC unfinished genomic DNA sequence of N. meningitidis. BASB024
CC polypeptides and polynucleotides are useful for generating an
CC immune response in an animal. Antibodies specific BASB024 polypeptides
CC are useful for treating N. meningitidis infection, which causes
CC bacteremia and meningitis.
XX
SQ Sequence 922 AA;
```

```
Query Match 97.5%; Score 4769.5; DB 21; Length 922;
Best Local Similarity 97.7%; Pred. No. 0;
Matches 901; Conservative 8; Mismatches 12; Indels 1; Gaps 1;
```

```
QY 1 MRSSFRLKPICFYLMGVTLHYHYAEADAGRAGSEAIQVLEDVHVHAKRVPKDKVFTDA 60
Db 1 mrssfrikpicfyimgvmlhyhhsyaedagragseaqvleedvhvkavpkdkkvftda 60
QY 61 RAVSTRQDIFKSSSENLDNIVRSIFCAFTQODKSSGIVSLNIRGDSGFRVNTWVDGIQT 120
Db 61 ravstrqdifkssennldnivrslpfaftqgdksgivslnirgdsgrvntwvmdgitqt 120
QY 121 FYSTSTADAGRAGGSOFGASVDSNFIAGLDVVYKGSFGSAGINSLAGSANLRTLGVDDVV 180
Db 121 fyststdagraggsqfgasvdsnfiaqldvvykgsfsgsaginslagsanrltlgvddvv 180
QY 181 QGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESASGVLYGHSRRSVAQNIRVGGG 240
Db 181 qgnntyglllkgltgtnstkgnamaaigarkwlesasgvlyghsrrsvaqnirvsgg 240
QY 241 QHIGNFGAEYLERRKORYFVQEGALKFNSDSGKWERDLQROQWKYPKYNKN-QELQKY 299
Db 241 qhignfgaeylerrrkoryfvqegalkfnsnsgkwerdfqrpywktyqkynndpqelqky 300
QY 300 IEGHDKSWRENLAPOYDITPIDPSLSKQOSAGNLFKLEYDGVFNKYTAQFRDLNFKIGSR 359
Db 301 ieghdkswrenlapqyditpidpslskqosagnlfkleydgvfnkytaqfrdlntkigsr 360
QY 360 KIINRNQFNYGLSLNPTNLNLTAAYNSGRQKYPKSGKFTGWLGLLKDFEYNNAKIIDL 419
Db 361 kiinrnyqfnyglslnsyanlnltaaynsgrqkypksgkftgwglldkfetynnakildl 420
QY 420 NNTATFRLPRETELQTLTGPNYFNEYCKNRPPEELGIFFDGPDQDNGLYSLGRFGDK 479
Db 421 nntatfrlpretelqtltgfnymeygnrffpeelgiffdgpddnglyslgrfgdk 480
QY 480 GLLPKQSTIVOPAGSQYFNTFYDAALKKDIYRLNLTNTVGYRFGGEYTYGYGSDDEFK 539
Db 481 glpkqstivopagsqyftntfydaalkkdiyrlnystntvgyrfggeytygygsddefk 540
QY 540 RAFGENSPYKKHCNRSRGYIPEVVLKYGKRRANNHVSISADFGDYPMPFASYSRTHRM 599
Db 541 rafgensptykhhcnqscgiyepvvlkyygkrrannhsvsisadfgdyfmpfasyrthrm 600
```

Qy 600 PNIQEMYSQIGDSGVHTALKPERANTWQFGNTYKKGLLKQDDTLGLKLGYRSRIDNY 659  
Db 601 pniqemyfsqigdsghvhtalkperantwqfgntyykkgllkqddtlglklgyrsridny 660  
Qy 660 IHNVYKWKWDLNGDIPSWSVSTGLAYTIQHRNFKDKVHKHGFELNLDYGRFTNLSYA 719  
Db 661 ihnvvgkwdlmgpwsvsstglaytiqhrnfkdvkhkgfelenldygrfftnlsya 720  
Qy 720 YOKSTQPTNFSDAESPNNASKEDQKQGYGLSRVSALPRDYGRLEVGTRWLGKNTLGG 779  
Db 721 yqkstqptnfsdaespnnaskedqlkqgyglrsvalprdygrlevgtwlgkntl19g 780  
Qy 780 AMRYFGKSIRATAERYIDGTNGNTSNFRLGKRSIKQTETLARQPLIFDYAAYEPKK 839  
Db 781 amryfgksirataeeryidgtngntsnvrqlgkrsikqtetlarqplifdyaayepkk 840  
Qy 840 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFDPKDDEYTCNADKTLGCKYGGT 899  
Db 841 nlifraevknlfdriryidpldagndaatqryssfdpkdkdeevtcnadtclngkyggt 900  
Qy 900 SKSVLTNFAGRFTLITMSYKF 921  
Db 901 sksvltnfargrtflitmsyxf 922

RESULT 4  
AAV38940  
ID AAV38940 standard; Protein; 922 AA.  
XX  
AC AAV38940;  
XX  
DT 08-OCT-1999 (first entry)  
XX  
DE N. gonorrhoeae antigen encoded by a variant ORF133.  
XX  
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.  
XX  
OS Neisseria gonorrhoeae.  
XX  
PN WO9924578-A2.  
XX  
PD 20-MAY-1999.  
XX  
XX  
PF 09-OCT-1998; 98WO-IB01665.  
XX  
PR 01-SEP-1998; 98GB-0019016.  
PR 06-NOV-1997; 97GB-0023516.  
PR 14-NOV-1997; 97GB-0024190.  
PR 18-NOV-1997; 97GB-0024386.  
PR 27-NOV-1997; 97GB-0025158.  
PR 10-DEC-1997; 97GB-0026147.  
PR 14-JAN-1998; 98GB-0000759.  
XX  
PA (CHIR-) CHIRON SPA.  
XX  
PI Grandi G, Masiagnani V, Pizza M, Rappuoli R, Scarlato V;  
XX  
DR WPI; 1999-327407/27.  
DR N-PSDB; AA212354.  
XX  
XX  
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
PT diagnosis, treatment and prevention of infection  
XX  
PS Claim 4; Page 480; 524pp; English.  
XX  
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis  
CC and N. gonorrhoeae antigenic proteins. They are encoded by open  
CC reading frames (ORFs) AA211972-212358. The antigenic proteins,  
CC their fragments, their nucleic acids and antibodies are used for  
CC diagnosis, prevention (as vaccines) or treatment of Neisseria  
CC infections, such as meningitis, septicaemia and gonorrhea. Both  
CC organisms are closely related. Fragments of the nucleic acids

CC are useful as hybridisation probes and antisense reagents.  
XX  
SQ Sequence 922 AA;  
Query Match 96.2%; Score 4707.5; DB 20; Length 922;  
Best Local Similarity 96.0%; Pred. No. 0;  
Matches 885; Conservative 15; Mismatches 21; Indels 1; Gaps 1;  
Qy 1 MRSSFRLLKPCFYLMGVTLHYHYAEADAGRAGSEAOQVLEDVHVAKRPVKDKKVTDA 60  
Db 1 mrsfrllkpcfylmgvmlhyhsyaedagragseaqvledvvhvakrpkdkkvtfda 60  
Qy 61 RAVSTRQDIFKSSNLDNIVRSIPGAFTQDDKSSGIVSLNIRGSGRGRVNTWVDGTQOT 120  
Db 61 ravstrqdvfksgenldnivrsipgaftqddkssgivslnirgsggrvntwvdgtqt 120  
Qy 121 FYSTDAGRAGSSQFGASVDSNFIAGLDVYKGSFSGSAGINSAGSANTLRTGVDDV 180  
Db 121 fystdtagragssqfgasvdsnfiaagldvvykgsfsgsaginsagsanltrtgvddv 180  
Qy 181 QGNNTYGLLLKGLTGTNSTKGNMAAATGARKWLESASVGLYCHSRSSVAQNYRVGGG 240  
Db 181 qgnntyglillkgltgtnstkgnmaaaigarkwlesgasvgylyghsrgrvaqnyrvggg 240  
Qy 241 QHIGNFGAELYERRKQRYFVOEGALKFNSDGSQKWERDLQROQWKYKPKYNN-OELQKY 299  
Db 241 qhignfgeeylerkkqyfvqegglknagsgkwerdlrqywktykkyedqelqky 300  
Qy 300 IEHDKSWRENLAPOYDITPIDPSSLKOQAGNLFKLEYDGVFNKYTAQPRDLNTKIGSR 359  
Db 301 ieehdkswrenlapqyditpidpsglkqgsagnlfkleydgvfnkytaqprdlntigrs 360  
Qy 360 KIINRNYQFNYSGLNPNYTNLNTAAYNNGRQKPKSGKFTGWCGLLKDFETYNNAKILDL 419  
Db 361 kiinrnyqfnyslnpytnlnitaaynsgrqypkagktgwgllkdfetynnakildl 420  
Qy 420 NNTATFLPRETELOTTLGFNFHNEYKKNRFPPELGLFFDGDQDNGLSYLRGRFGDK 479  
Db 421 nntatflpretelqtltgfnfhneygknrfpeelglffdgpdqnglsyigrfgdk 480  
Qy 480 GLLPQKSTIVQAGSYFNTFYDAAALKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFK 539  
Db 481 glipqkstivqagsyfnfyfdaalkdiyrlnysntvgyrfggeytgyygsenefk 540  
Qy 540 RAFGENSPYKKHCNRSCGIYEPVLKKYKGRANNHVSISADGDFEMPEFASYSRTHM 599  
Db 541 rafgenspaykehcdpscglyepvlkkygkkrannhvsisadgdfmpfagysrthm 600  
Qy 600 PNIQEMYFSQIGDSGVHTALKPERANTWQFGNTYKKGLLKQDDTLGLKLGYRSRIDNY 659  
Db 601 pniqemyfsqigdsghvhtalkperantwqfgntyykkgllkqddtlglklgyrsridny 660  
Qy 660 IHNVYKWKWDLNGDIPSWSVSTGLAYTIQHRNFKDKVHKHGFELNLDYGRFTNLSYA 719  
Db 661 ihnvvgkwdlmgpwsvsstglaytiqhrnfkdvkhkgfelenldygrfftnlsya 720  
Qy 720 YOKSTQPTNFSDAESPNNASKEDQKQGYGLSRVSALPRDYGRLEVGTRWLGKNTLGG 779  
Db 721 yqkstqptnfsdaespnnaskedqlkqgyglrsvalprdygrlevgtwlgkntl19g 780  
Qy 780 AMRYFGKSIRATAERYIDGTNGNTSNFRLGKRSIKQTETLARQPLIFDYAAYEPKK 839  
Db 781 amryfgksirataeeryidgtngntsnvrqlgkrsikqtetlarqplifdyaayepkk 840  
Qy 840 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFDPKDDEYTCNADKTLGCKYGGT 899  
Db 841 nlifraevknlfdriryidpldagndaatqryssfdpkdkdeevtcnadtclngkyggt 900  
Qy 900 SKSVLTNFAGRFTLITMSYKF 921  
Db 901 sksvltnfargrtflitmsyxf 922

```
RESULT 5
AAV38937
ID AAY38937 standard; Protein; 888 AA.
XX
AC AAY38937;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria meningitidis antigen encoded by a partial ORF133.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria meningitidis.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR WP1; 1999-327407/27.
DR N-PSDB; AA212352.
XX
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
PS Claim 4; Page 474; 524pp; English.
XX
CC Amino acid sequences AAY38499-X38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AA211972-212358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
SQ Sequence 888 AA;

Query Match 96.1%; Score 4703; DB 20; Length 888;
Best Local Similarity 99.7%; Pred. No. 0;
Matches 885; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

QY 34 EAQIQVLEDVHVAKRVPKDKVFTDARAVSTRQDIFKSSSENLDNIVRSTPGAFTOQDKS 93
D 1 eaqiqlvedvhvakrvpkdkvftdaravstrqdfksssenldnivrslpgaftqdkds 60
QY 94 SGIVSLNIRGDSGFRVNTMWVDGITQTFYSTSTDAGRAGSSQFGASVDSNFIAGLDVVK 153
D 61 sgivslnirgdsgrvntmvdgitqtfyststdagragssqfgasvdsnfiagldvkv 120
QY 154 GSPFSGAGINSLAGSANLRTGLGVDDVVQGNNTYGLLLKGLTGNTSKGNAMAAIGARKWL 213
D 121 gsfsgaginslagsanlrltglgvddvvqgnntyglllkglgtgntskgnamaaigarkwl 180
QY 214 ESGASGVLYGHSSRRSVAQNYRVGGGGQHICNFGAEYLERRKORYFVOEGALKFNSDSGK 273
D 181 esgasgvlyghssrrsvaqnyrvvgggqhignfgaeylerkroryfvoegalkfnsdsgk 240
```

```
QY 274 WERDLQROQWKYPKYKNYNNOELQKYIEGHDKSWRENLAPOYDITPIDPSLKOQSAGNL 333
D 241 werdlqrqwkypkyknynnqelqkyieehdkswrenlpxpyditpidpslkgqsagnl 300
QY 334 FKLEYDGVFNKYTAQFRLNTKIGSRKIINRNYQFNGLSNPYTNLNLTAAYNSGRQKY 393
D 301 fkleydgvfnkytaqfrrlntkigsrkilnrnyqfnvglslnpytnlnltaaynsgrqky 360
QY 394 PGSKFTGWLKGLKDFETYNNAKILDLNNTATFRLPRETELTQTTLGFGNYFHNEYKRNFFE 453
D 361 pgskfthgwlkglkdfetynnakildlnntatfrrlpretelqtltglfnyfhneykrfpe 420
QY 454 ELGLFFDPDQDNGLYSLGRFKGDKGLLPKSTIVOPAGSQYFNTFYDAALKKDIYRL 513
D 421 elglffdpdqdnnglyslgrfkdgkllpqkstivpagsqyfnfyfdaalkkdiyrl 480
QY 514 NYSTNTVYRFGGEYTGYYGSDDEFFKRAFGENSPYKHKHNRSCGIYEPVLKYYKKRAN 573
D 481 nystntvyrfggeytgyygdsddefkrafgensptykhhnrscgiyepvlkyykkran 540
QY 574 NHSVTSADFCDFMPFPASYSRTHRMPIQEMFYFSQIGDSGVHTALKPERANTWQFGFNT 633
D 541 nhsvtsadfcdfmpfpasysrthrmpiqemyfsqigdsgvhtalkperantwqfgfnt 600
QY 634 YKGLLKQDDTLGLKLVGYRSRIDNYIHNVYKWKWDLUNGDIPIWSVSTGLAYTIOHRNFK 693
D 601 ykglkkddtllglkvlgyrsridnyihnvyygkwwdlungdipisvstglaytiqhrnfk 660
QY 694 DKVHKGFELELVNDYGRFETNLSYAYOKSTOPTNFSDASESPNNASKEDQLKQGYGLSR 753
D 661 dkvhkgfelelvndygrfettlnlsyayokstoptnfsdasespnnaskeqqlkgyglsr 720
QY 754 VSALPRDYGRLEVGTWRWLGKLTGGAMRYFGKSIRATAEERYIDGTNGGNTSNFRLGK 813
D 721 vsalprdygrlevgtwrwlgkltggamryfgksirataeeryidgtnggntsnfrlgk 780
QY 814 RSKQETELARQPLIFDFYAAEPKKNLIFRAEVKNLFDPRYIDPLDAGNDAATQRYSS 873
D 781 rskqetelarqp lifdfyaaepkknli fraevknlfdr yidpldagndaatqryyss 840
QY 874 FDPKDKDEDVTCNADKTLCKNGKYGSTKSVLTNPFARGRTFLITMSYKF 921
D 841 fdpkdkdedvtcnadkltclngkygstsksvltnfargrtflmtmsyxf 888

RESULT 6
AAV38939
ID AAY38939 standard; Protein; 922 AA.
XX
AC AAY38939;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria gonorrhoeae antigen encoded by ORF133.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria gonorrhoeae.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
```

PR	14-JAN-1998;	98GB-0000759.	
XX	(CHIR-) CHIRON SPA.		
XX			
PI	Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;		
XX	WPI; 1999-327407/27.		
DR			
XX			
PT	Proteins from <i>Neisseria meningitidis</i> and <i>N. gonorrhoeae</i> useful for		
PT	diagnosis, treatment and prevention of infection		
PS	Claim 4; Page 479; 524pp; English.		
XX			
CC	Amino acid sequences AAY38499-Y38944 represent <i>Neisseria meningitidis</i>		
CC	and <i>N. gonorrhoeae</i> antigenic proteins. They are encoded by open		
CC	reading frames (ORFs) AA211972-212358. The antigenic proteins,		
CC	their fragments, their nucleic acids and antibodies are used for		
CC	diagnosis, prevention (as vaccines) or treatment of <i>Neisseria</i>		
CC	infections, such as meningitis, septicaemia and gonorrhea. Both		
CC	organisms are closely related. Fragments of the nucleic acids		
CC	are useful as hybridisation probes and antisense reagents.		
XX			
SQ	Sequence :922 AA;		
	Query Match 96.0%; Score 4696.5; DB 20; Length 922;		
	Best Local Similarity' 95.8%; Pred. No. 0;		
	Matches 883; Conservative 15; Mismatches 23; Indels 1; Gaps 1;		
QY	1 MRSSFRLKPTCFYLMGVTLVHYSVAEDAGRAGSEAOIQVLEDVHVAKRPDKKVFPTDA 60		
DB	1 mrsfrrlkptcfylmgvmlhnhysvaedagragseadqlvledvhvakrvpdkkvftda 60		
QY	61 RAVSTRODIPFKSENLDNIVRSIPGAFQTQODKSSGVLSNIRGDSGFGRYNTMVDGITQT 120		
DB	61 ravstrdvfksgenldnivrspgafqtqgdkssgvlslnirgdsfggrvntmvdgitqt 120		
QY	121 FYSTSTDAGRAGSSQPGASVDNSFTAGLDVVGKSGSGAGINSLAGSANLRTLGVDDVV 180		
DB	121 fyststdagragssqpgasvdnsftagldvvgksgsgaginslagsanlrtlglvddvv 180		
QY	181 QGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLIESGASGVLYGHRSRVAQNYRVGGGG 240		
DB	181 qgnntyglllkgltgtnstkgnamaaigarkwlesgasgvlyghrsrvaqnyrvvgggg 240		
QY	241 QHTGNFGAEYLERRKORYFVOEGALKFNSDSGKWERDLORQOWKYPKYNYNH-OELQKY 299		
DB	241 qhignfgeylelrkqgyfvqegglkfnagsgkwerdlgrqywktykkyedpqelqky 300		
QY	300 IEGHDKSWRENLAPOYDITPIDSSLKQOSAGNLFKLEVDGVFNKYTAQFRDLNTKIGSR 359		
DB	301 iehdkswrenlapoyditypidpsglkqgsagnllfnleydgvfnkytaqfrdlntrigr 360		
QY	360 KIINRNYQFNYGSLNAPYTNLNLTAAYNSGRQKYPKSGFTGWGLKDPETYNNAKILDL 419		
DB	361 kiinrnyqfngyslapytnlnltaaynsgrqkypkgakftgwgllkdfetynnakild. 420		
QY	420 NWTATFRLPRETELQTTFLGNYPHNEYKGNRPPEELGLFFDQDQONGLYSYLGRFKGDK 479		
DB	421 nwtatfrlpretelqttltnfyfhneyknrfpeelglffdpqngnlysyldrfgdk 480		
QY	480 GLLPQKSTIVOPAGSYFNTFYDAAALKDDIVRLNLTNTVTGYRFGGEYTGYYGSDDEFK 539		
DB	481 glldpqsstivopagsyfnfcfydlaalkddivrlnltntvtnfyrggeytygysenefk 540		
QY	540 RAFGENSPYKKHCNRCSGYIEPVLKKYKGRANNHVSISADFGDYFMPFASYSRTHRM 599		
DB	541 rafgenspaykehcdpscglyepylkkykgrannhsvsisadfgdyfmpfagsyrthrm 600		
QY	600 PNTQEMFYSGISGSHVHTALKPBRANTWFGPNYKGLLKQDDTGLGLKLVGYRSRIDNY 659		
DB	601 pntqemfysgisgshvhtalkpbrantwfgfntykgllkqddilglklvgyrsridny 660		

QY	278	LQRQQ-----W--KYK-PYKNY-----NNQELQKYIEGHDKSWRENLAQYDITPID	321
QY	278	LQRQQ-----W--KYK-PYKNY-----NNQELQKYIEGHDKSWRENLAQYDITPID	321
DB	287	l tkldketgkplwdryqyfggkcyglcldtkekfdveyadvqqqkgakgkeyslatpdl	346
QY	322	PSSLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRKIIIRNTQFNYYGLSLNPTYNLN	381
DB	347	italnqtkskhlakirvnnndtsdvalqlrkmdtttqsrriisndvqldaaaynpneidlk	406

Query Match 52.9%; Score 2568.5; DB 21; Length 947;  
Best Local Similarity 55.6%; Pred. No. 7.4e-179;  
Matches 503; Conservative 144; Mismatches 232; Indels 25; Gaps 12;

Qy	39	VLEDVHVYKAKR-vpKDKKVTFDARAVYSTRODIFKSESENLDNIVRSITPGAFATQODKSSGIV	97
Db	48	ildevvvtatngtkksqkptkasatsvrenfnasenidaiVrsypgafatqgdkssgiv	107
Qy	98	SLNIIRGDSGRCRVNTWVDGJTQTFYSTSTDAGRAGSSQFGASVDSNFAGTGLDVKVGSFS	157
Db	108	slnvrgdsgransmdvgvtqtfystsdagrggtsqfavidqnlafagvelnkgsgfn	167
Qy	158	GSAGINSLAGSANLRTIGVDVQGNNTYGLLLKGLTGTNSTKGNMAAIGARKWLESGA	217
Db	168	gkgsinlittsantfrtlnaadvladdknfgiaiktgknatdknfmlaagrgfwldng-	226
Qy	218	SVGLYIGHSRRSVAQNYRVGGGQHIGNFGAEYLERRKORYFVQEGALAFNSDSGKWERD	277
Db	227	sisalayshkdisqnykvvgggtghnvgvddlllsekqkvfakelhalynearswqkd	286
Qy	278	LQROQ-----w--KYK-PYKN-----NNLOLQKYIEGHDKSENLAIPQYDITPT	321
Db	287	ltkldketgplwdrkyqfggkcygIgcldtkekfdeyvadkqqqwkghgakeysitpid	346
Qy	322	PSSLKQGSAGNLKFLKYVDGVFNXYTAQFRDLNTKIGSRKIINRYQFNYGLSNPYTNLN	381
Db	347	italnatskshakirvymdtsdvqIqrkmdttisrrisndvnyglvaaynpnneidlk	406



Db 565 epilhsgkhkafnhsatlssaelssdyfmpfftystrhrmpnigemffsfqvsngvntalk 624  
 QY 621 PERANTWQFGFNYYKKGLLKQDDTLGLKLVGYRSRDINYLHNYGKWDNLNGDIPSWSS 680  
 Db 625 pedsdyqlgintytkgltqddvlgklygysfknynhnygyvw--rdgmptwaes 682  
 QY 681 TGLAYTIOHRNFKDKVHKHGFELNLDYGRFTNLSYAYOKSTQPTNFSDESPPNAS 740  
 Db 683 ngfkytiiahnykpvkksveleindymgrffanvsyayqrtnqpcnycadasprnnas 742  
 QY 741 KEDOLKOGYGLSRVSLPRDYGRLEVRWLGKLTGLGAMRYFGKSRATAEERYIDGT 800  
 Db 743 qedilkgyglslrsvmlpkdygrlelgrwfdqkltlglaaaryygskratieeeyings 802  
 QY 801 NGGNTSFRLGKRSIKQTETLARQPLIFDYAAYPEKKNLIFRAEVKNLFDRIYDPLD 860  
 Db 803 r-fkntllrrenyavvktedikkpildlhvsyepikdliikaevgnlildkryvdpld 861  
 QY 861 AGNDAAATORYSSFDPKDDEYTCNADKTLGCKGYGTSKSVLTNFARGRTFLITMSYK 920  
 Db 862 agndaasqryssl-----nnsiecaqdsac----ggsdktvlynfargrtvylisnyk 912  
 QY 921 F 921  
 Db 913 f 913  
 RESULT 10  
 AAY94672  
 ID AAY94672 standard; Protein; 918 AA.  
 XX  
 AC AAY94672;  
 XX  
 DT 01-DEC-2000 (first entry)  
 XX  
 DE Haemophilus antigen BASB070 protein sequence.  
 XX  
 KW Vaccine; BASB070; Haemophilus influenzae; strain nH3224; pneumonia;  
 KW chronic bronchitis; sinusitis; otitis media; meningitis; antigen;  
 KW systemic disease; outer membrane protein.  
 XX  
 OS Haemophilus influenzae.  
 XX  
 PN WC200050599-A1.  
 XX  
 PD 31-AUG-2000.  
 XX  
 PF 22-FEB-2000; 2000WO-EP01423.  
 XX  
 PR 24-FEB-1999; 99GB-0004183.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 PI Ruelle J, Thonnard J;  
 XX  
 DR WPI; 2000-543912/49.  
 DR N-PSDB; AAA28031.  
 XX  
 PT Vaccinating against Haemophilus influenzae using BASB070 polypeptides  
 XX and/or the nucleic acids that encode them -  
 PS Claim 1; Fig 2; 97pp; English.  
 XX  
 CC This invention relates to a vaccine composition which contains BASB070  
 CC polypeptides from Haemophilus influenzae and/or the nucleotide sequences  
 CC that encode them. BASB070 DNA sequences encode proteins which are  
 CC integral outer membrane proteins with a beta-barrel conformation. The  
 CC invention includes BASB070 nucleotide and protein sequences from  
 CC H. influenzae strains RdKw20 and nH3224. The invention includes an  
 CC expression vector and recombinant live microorganism comprising a BASB070  
 CC polynucleotide sequence, a process for the production of a BASB070  
 CC protein, an antibody specific for the two BASB070 proteins of the  
 CC invention, and a method for diagnosing an H. influenzae infection.

CC Haemophilus influenzae is a common cause of pneumonia, exacerbation of  
 CC chronic bronchitis, sinusitis and otitis media, and H. influenzae type b  
 CC causes bacterial meningitis and systemic diseases. The vaccine exhibits  
 CC antibacterial activity, and is used in the preparation of an agent for  
 CC use in generating an immunological response in a mammal.  
 CC The present sequence represents a BASB070 protein isolated from  
 CC H. influenzae strain nH3224. The protein is used in the production of  
 CC the vaccine of the invention.  
 XX  
 XX Sequence 918 AA;  
 QY  
 Query Match 50.7%; Score 2483.5; DB 21; Length 918;  
 Best Local Similarity 51.1%; Pred. No. 2.9e-171;  
 Matches 494; Conservative 154; Mismatches 225; Indels 93; Gaps 17;  
 QY 1 MRSFRLKPCIFYL---MGVTLHYHYAEADAGRAGSQAQIQVLEDVHVAKARVPDKKVF 57  
 Db 1 mkkaiklnltlsintigmtitqaqaetlg-----qidvvekv-----isndkkpf 48  
 QY 58 TDARAVTRQDIFKSENLDNIVRSIFGAFQQDKSSGIVSLNIRGDSGFRVNTWVDGI 117  
 Db 49 teakakstrenvfkettidqvirsipgafqgdkgsvsvnirgenglgrvntmvdgv 108  
 QY 118 TQTFYSTDAGRAGSSQFGASVDSNFIAGLDVYKGSFSGSAGINSIAGSANLRTLGVD 177  
 Db 109 tqtfystaldsgsgsgsgfgaaidpnfiagvdknfnsgaginalagalsanrtlsn 168  
 QY 178 DVQGNNTYGLLLKGLTGTNSTKGNMAAICARKWLSGASGVLYGHRSRVSQAQNYRVG 237  
 Db 169 dvitddkpfgiikgmtsgnatksnfmttaagrkwldnggyvgyvgyqsrevsqdyri- 227  
 QY 238 GGGQHGICFCAEYLERRKQRYFVQEGALKFNSDGKWERDLQROOW----- 283  
 Db 228 gggerlaslsggdilakekelfrindgyv-lns-agqwapdlnkphwscntpslkdksms 285  
 QY 284 -KYPYK-----NYNQELQY---IEGHDKSWRENLAPOYDIT 318  
 Db 286 tsckpyrlgpaatrteqllkelledgkpdiekqlsgndgieeteksfern-kdydva 344  
 QY 319 PIDPSSILKQOSAGNLKFLKLEYDVFNKYTAQFRDLNTIGSKRIINRYQFNYGSLNPLY 378  
 Db 345 piepgslqsrshrshllkfeysddhhtlgaqirtldnkigsrkienryqynfnnsyl 404  
 QY 379 NLNLTAAYNSGRQYKPGSKFTGWGLLKDFETYNNAKLLDNTATRLPRETELQTTLG 438  
 Db 405 dlnlmaahnigktyipkgffagvgvadklltknvanivdinnshtflpkeidlkttlg 464  
 QY 439 FNYFHEYGNRNPPEELGLPEFDGPDQDNGLYSY--LGRFKDGKGLLPQKSTIVOPAGSOY 496  
 Db 465 fnyftneysknrfpeelsllyvneshdqgylslsnkgrysgskglipqrsvillqpsgkqk 524  
 QY 497 FNTFYDAALKKDIYRLNYSYNTNTVTYRFGGTYGYYSDDDEFKRAFGENSPYTKKHCNRS 556  
 Db 525 fktvyfdalskgyihlnysvntfhyafngyvykntadk----- 565  
 QY 557 CGIYEPVLKXYGKKRANNSVSIADFGDPMFPASYSRTHRMNIOEMYSQDGSVH 616  
 Db 566 --lnepilhsgkhkafnhsatlssaelssdyfmpfftystrhrmpnigemffsfqvsdavn 623  
 QY 617 TALKPERANTWQFGFNYYKKGLLKQDDTLGLKLVGYRSRDINYLHNYGKWDNLNGDIPS 676  
 Db 624 talkpedsdyqlgintytkgltqddvlgklygysfknynhnygyd-wsrdgvmpe 682  
 QY 677 WVSSTGLAYTIOHRNFKDKVHKHGFELNLDYGRFTNLSYAYOKSTQPTNFSDESPP 736  
 Db 683 warlingfrltiahnykpvkksveleindymgrffanvsyayqrtnqpcnycadasarp 742  
 QY 737 NNASKEDQLKQGYGLSRVSLPRDYGRLEVRWLGKLTGLGAMRYFGKSRATAEERY 796  
 Db 743 rnaskeeilkggyglslrsmplpkdygrlelgrwfdqkltlglaaaryygskrattgeey 802  
 QY 797 IDGTN-CGNTSNFRLGKRSIKQTETLARQPLIFDYAAYPEKKNLIFRAEVKNLFDRIY 855

```

      803  ingryeknttrdriy--yaikteeikkpfiidllhvsyepikdliikaevqalldkry 860
      856  IDPLDAGNDAATORYYSSFDPKDDEVTNADKTLGNGYGGTSKSVLTNFAARGTFLI 915
      861  vdpldagndaasqrryssl-----ndslackinestcn---dsektivlynfargrtyil 912
      916  TMSYKF 921
      913  slnykf 918

RESULT 11
AA38936
ID AAY38936 standard; Protein; 393 AA.
XX
AC AAY38936;
XX
DT 08-OCT-1999 (first entry)
XX
DE Neisseria meningitidis antigen encoded by a partial ORF133.
XX
KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;
KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.
XX
OS Neisseria meningitidis.
XX
PN WO9924578-A2.
XX
PD 20-MAY-1999.
XX
PF 09-OCT-1998; 98WO-IB01665.
XX
PR 01-SEP-1998; 98GB-0019016.
PR 06-NOV-1997; 97GB-0023516.
PR 14-NOV-1997; 97GB-0024190.
PR 18-NOV-1997; 97GB-0024386.
PR 27-NOV-1997; 97GB-0025158.
PR 10-DEC-1997; 97GB-0026147.
PR 14-JAN-1998; 98GB-0000759.
XX
PA (CHIR-) CHIRON SPA.
XX
PI Grandi G, Massignani V, Pizza M, Rappuoli R, Scarlato V;
XX
DR WPI; 1999-327407/27.
DR N-PSDB; AAZ12351.
XX
PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for
PT diagnosis, treatment and prevention of infection
XX
PS Claim 4; Page 473; 524pp; English.
XX
CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis
CC and N. gonorrhoeae antigenic proteins. They are encoded by open
CC reading frames (ORFs) AAZ11972-Z12358. The antigenic proteins,
CC their fragments, their nucleic acids and antibodies are used for
CC diagnosis, prevention (as vaccines) or treatment of Neisseria
CC infections, such as meningitis, septicaemia and gonorrhea. Both
CC organisms are closely related. Fragments of the nucleic acids
CC are useful as hybridisation probes and antisense reagents.
XX
SQ Sequence 393 AA;

Query Match 40.8%; Score 1996; DB 20; Length 393;
Best Local Similarity 95.2%; Pred. No. 1.9e-136;
Matches 373; Conservative 3; Mismatches 16; Indels 0; Gaps 0;

QY 530 GYGSDDDFKRAFGENSPYKHKHCRSCGIYEPVLKKYKKRANHSVSISADFGDYFMP 589
      |||||
Db 2 gyygsdddfkrafgensptkxkhcnrscgiyepvlkkykkranhsvsisadfgdyfmp 61

```

```

QY 590 FASYSTRHRMPNIQEMYFSQIGDSGVHTALKPERANTWQCFNTYKKGLLKQDDTLGLKL 649
      |||||
Db 62 fasystrhrmpniqemyfsqigdsgvhtalkperantwqcfntykgllkqddtlglkl 121
      |||||
QY 650 VGRSRIDNTIHNHYGKWDNLNGDIPSWSVSTGLAYTIOHRNFKDKVHKHGFLELNDY 709
      |||||
Db 122 vgyrsridnyihnygkwdlngdipswsvstglaytiqhrxfkdkvhqxxxxxxxydy 181
      |||||
QY 710 GRPFTNLISYAVOKSTOPTNFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGT 769
      |||||
Db 182 grftnlisyayqkstqptnfsdasespnnaskedqlkqgyglrsvalprdygrlevgt 241
      |||||
QY 770 WLGNKLTGGMARYFGKSIRATAEERYIDGTNGGNTSNFRQLGRSIRKQTTETLARQPLIF 829
      |||||
Db 242 wlgnkltlgamaryfgksirataeeryidgtnggntsnfrqlgrsikqetlarqplix 301
      |||||
QY 830 DFYAAEPKKNLIFRAEVKNLFDPRYIDPLDAGNDAATORYYSSFDPKDDEVTCTNADK 889
      |||||
Db 302 dfnaayepkknlfraevknlfdrdyidpldagndaatoryyssfdpkdkdxdtcnadk 361
      |||||
QY 890 TLCNGKYGGTSKSVLTNFAARGTFLITMSYKF 921
      |||||
Db 362 tlcngkyggtsksvltnfargrtflitmsykf 393
      |||||

RESULT 12
ABB52934
ID ABB52934 standard; Protein; 753 AA.
XX
AC ABB52934;
XX
DT 11-FEB-2002 (first entry)
XX
DE Escherichia coli polypeptide SEQ ID NO 1277.
XX
KW Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW systemic infection; non-diarrhoeal infection; septicaemia;
KW pyelonephritis; antibiotic resistance.
XX
OS Escherichia coli.
XX
PN WO200166572-A2.
XX
PD 13-SEP-2001.
XX
PF 12-MAR-2001; 2001WO-EP03445.
XX
PR 10-MAR-2000; 2000FR-0003145.
PR 02-FEB-2001; 2001FR-0001449.
XX
PA (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
XX
PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
XX
DR WPI; 2001-550253/61.
XX
PT A library of DNA fragments of Escherichia coli strains for the
PT phylogenic determination of a given strain comprises polynucleotides of
PT nature B2/D+ A- -
XX
PS Example 6; Fig 6; 646pp; English.
XX
CC The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC and encoded proteins (ABB52459-ABB52919) of nature B2/D+A-. The
CC polynucleotides have potential antiinflammatory, antibacterial and
CC immunosuppressive activity as part of pharmaceutical compositions used to
CC treat, palliate or prevent extra-intestinal E. coli infections. The
CC polypeptides are useful for determining the phylogenic group of a given
CC E. coli strain. These polypeptides can detect and treat an undesired
CC development of E. coli, particularly an extra-intestinal infection that
CC include systemic and non-diarrhoeal infections such as septicaemia,

```



pyelonephritis and meningitis this is particularly advantageous as bacterial resistance is increasing with the more frequent use of broad spectrum antibiotics.

xx Sequence 753 AA:

sq

Query Match	18.3%	Score 898;	DB 22;	Length 753;
Best Local Similarity	28.7%	Pred. No. 1.9e-56;		
Matches 262;	Conservative 118;	Mismatches 296;	Indels 236;	Gaps 26;

Qy	45	VKARVPDKKVFYTDARAVSTRQDIFPKSSSENLDNIVRSPGATFOQDKSSGIVSLNIRGD	101
Db	43	vsvgtktseaealektgatsr--ttkdnlgslatvrsmptgtytqldpqqgalsvairgm	101
Qy	105	SGFGRVMTWGDITQTFYSTSDAGRAGGS--SOFGASVDSNFIAGLDVYVKGFSGSAGI	162
Db	102	sgfrvntmvdgltcqsfygtstgthstnmgavldpnlhavdvtrgdsgsgegl	161
Qy	163	NSLAGSANTLRGLVDDVVQGNNTYGLLLKGLTGCTNSTKGNAAMAIARK--WLESASGV	220
Db	162	nalgagsnmrtigvddvifngnyfgrsrfsvsgnslgrsgmialggksdafdtg--slg	220
Qy	221	VLYGHSRRSAQNYRVGGCGQHIGNFGAEYLERRKQRYFVQEGALKFNSDSCKWERDLQ	280
Db	221	vmaavsgsvsyntfsngsg-----	239
Qy	281	QOWKYKPYKNYNQOELQYIEGHDKSWRENLAPOYDITPIDPSLKKQSAGNLFKLEY-D	339
Db	240	-----inskef-----gydk-----ymqnpksqlykmdirp	266
Qy	340	GVFNKYTAQFRDLNWKGSRKIIINRYQFNWGLSLNPYTNL---NLTAAYNSGRQYXPKG	396
Db	267	defnsfelsartyenkfrditsddvlyky--hytpfseidfntvastgrnqkyrdg	324
Qy	397	SKFTGWLKDDFETYNNAKILDNNTATFPLPRETELOTTGLFNGYFHNEYKGNRFPPELG	456
Db	325	slyfyf---ktsaqrnsdaldnntsrftv-adndlefmlgsklmrty-----	369
Qy	457	LFFDGPDDNGLYSYLGRFGKDGKGLLPQKSTIVQPAGSYQNFYFDAAALKKDI---YR	512
Db	370	-----drtihsaagdpkangesi--ennpfapsqgqdisalytqlkvtrgiweadfn	419
Qy	513	LNYSNTVY-GVRFGEY-----TGYGSDDEKFRAGENSPTYKKHCNRCGIYEPVLK	565
Db	420	lnytrnitykpacdservicvpqgssydlcd-----	450
Qy	566	KYGKRRANNHSVISADFGDYFWPFASYSTRHMPNIQEMFYFSQIGDSGVHTALKPERAN	625
Db	451	---kegfnpsvqisaqvtpwlpqfigysksmrapniqemfnfnsagsmnpflkperae	507
Qy	626	TWQGFNTYKGLLKQDDTGLKLKLVGRSRIDNYIHN-----VYCKWWD	669
Db	508	twqgafndtrdliveqdarlfkalayrsiqnyisesylvcsgrgkcslpbevimgwe	567
Qy	670	LNGDIPSWVSSTGLAY*IQHRNFKDKVHKHGFLELNYDYGREFTNLVAYOKSTOPTNF	729
Db	568	---gisdeysdmlyyv-----nasdvialkgelemdydagfagrslsfsqqicdpts	620
Qy	730	SDASESPNNAKQDLKQGYGLRVSALPRDYGRLEVGTGRWLGNKTLTGAMAYFGKSI	789
Db	621	asth-----fgagditelpkymtldtgvrfdnaltgltkiyktgkarr	665
Qy	790	ATAERYIDGTNGNTSNFRLGKRSIKQPETLARQPLIFDFVAAAYEPKKNLIFRAEVN	849
Db	666	lspfedqentga-----likq--dlpqtiidlygtveyfmrnlklklsvgn	711
Qy	850	LFDRRYIDPLDAGNDAATQRYSSFDPKDDEVDTCNADKTLNCGYGGTSKSVLTNFAR	909
Db	712	lmnrdyseainkln-----mmpglgldethpans-----ar	741
Qy	910	GRTELTITMSYKF	921

Db 742 grtwifggdirf 753

RESULT 13  
ABB52968  
ID ABB52968 standard; Protein; 753 AA.

AC	ABB52968;
XX	
DT	11-FEB-2002 (first entry)
XX	
DE	Escherichia coli polypeptide SEO ID NO 1277.

AA	Escherichia coli; B2/Dr-A; anti-inflammatory; antibacterial;
KW	immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW	systemic infection; non-diarrhoeal infection; septicaemia;
KW	pylonephritis; antibiotic resistance.

OS *Escherichia coli*.

PN WO200166572-A2.

PD 13-SEP-2001.

XX  
PF 12-MAR-2001: 2001WO-EP03445.

XX  
PP 10-MAB-2000: 2000EB-0003115

PR 02-FEB-2001; 2001FR-0001449.

XX  
PA (INRM) INSERM INST NAT SANTE & RECH MEDICALE.

XX  
PI Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;

XX  
DP  
WPT: 2001-550253/61

XX A library of DNA fragments of *Escherichia coli* strains for the  
PT phylogenetic determination of a given strain comprises polynucleotides of  
PT nature B2/D+ A- -

PS Example 6; Fig 6; 646pp; English.

XX The invention relates to a library of DNA fragments of *Escherichia coli*  
CC strains comprising polynucleotides (ABBA88577-ABBA88729 and ABAA95933),  
CC and encoded proteins (ABB52459-ABBS52919 and ABB52954-ABBS53094) of nature  
CC B2/D+A-. The polynucleotides have potential antiinflammatory,  
CC antibacterial and immunosuppressive activity as part of pharmaceutical  
CC compositions used to treat, palliate or prevent extra-intestinal *E. coli*  
CC infections. The polypeptides are useful for determining the phylogenetic  
CC group of a given *E. coli* strain. These polypeptides can detect and treat  
CC an undesired development of *E. coli*, particularly an extra-intestinal  
CC infection that include systemic and non-diarrhoeal infections such as  
CC septicæmia, pyelonephritis and meningitis this is particularly  
CC advantageous as bacterial resistance is increasing with the more  
CC frequent use of broad spectrum antibiotics.

AA	Sequence	753 AA;
SQ		

Query Match	18.3%	Score 898;	DB 22;	Length 753;
Best Local Similarity	28.7%;	pred. NO.	1.9e-56;	
Matches 262;	Conservative 118;	Mismatches 296;	Indels 236;	Gaps 26;

**QY**      45   VKAKRVPKDKKVFTDARAVSTRQIDFKSSENLDNIVRSIPGAFTQDKSSGIVSLNRGD   104  
             | : ::    | | :    | : :    ||| : | : |    | : | : | : |  
**Dd**      43   vsvgkttseqalektgatssr-ttdknqlslatvrsmpgytytdpgqqaivsnirgm   101

```

QY 105 SGFGRVNTMDGITOTPTYSTTDAGRGS--SQFGASVDNSNFIAGLDVVKGSFSGSAGI 162
    |||||
nB 102 sgfgrvntmvdgitqsfvgrststgtttbqstnmgavlfdnpllvaydyvtrodssnsagi 161
    |||||

```

102 sfglrvltmuvqfqtzifgscgcttngscmmagvrtpphrrvavvrtgussgocg1 103  
104  
105  
106  
107  
108  
109  
110  
111  
112  
113  
114  
115  
116  
117  
118  
119  
120  
121  
122  
123  
124  
125  
126  
127  
128  
129  
130  
131  
132  
133  
134  
135  
136  
137  
138  
139  
140  
141  
142  
143  
144  
145  
146  
147  
148  
149  
150  
151  
152  
153  
154  
155  
156  
157  
158  
159  
160  
161  
162  
163  
164  
165  
166  
167  
168  
169  
170  
171  
172  
173  
174  
175  
176  
177  
178  
179  
180  
181  
182  
183  
184  
185  
186  
187  
188  
189  
190  
191  
192  
193  
194  
195  
196  
197  
198  
199  
200  
201  
202  
203  
204  
205  
206  
207  
208  
209  
210  
211  
212  
213  
214  
215  
216  
217  
218  
219  
220  
221  
222  
223  
224  
225  
226  
227  
228  
229  
230  
231  
232  
233  
234  
235  
236  
237  
238  
239  
240  
241  
242  
243  
244  
245  
246  
247  
248  
249  
250  
251  
252  
253  
254  
255  
256  
257  
258  
259  
260  
261  
262  
263  
264  
265  
266  
267  
268  
269  
270  
271  
272  
273  
274  
275  
276  
277  
278  
279  
280  
281  
282  
283  
284  
285  
286  
287  
288  
289  
290  
291  
292  
293  
294  
295  
296  
297  
298  
299  
300  
301  
302  
303  
304  
305  
306  
307  
308  
309  
310  
311  
312  
313  
314  
315  
316  
317  
318  
319  
320  
321  
322  
323  
324  
325  
326  
327  
328  
329  
330  
331  
332  
333  
334  
335  
336  
337  
338  
339  
340  
341  
342  
343  
344  
345  
346  
347  
348  
349  
350  
351  
352  
353  
354  
355  
356  
357  
358  
359  
360  
361  
362  
363  
364  
365  
366  
367  
368  
369  
370  
371  
372  
373  
374  
375  
376  
377  
378  
379  
380  
381  
382  
383  
384  
385  
386  
387  
388  
389  
390  
391  
392  
393  
394  
395  
396  
397  
398  
399  
400  
401  
402  
403  
404  
405  
406  
407  
408  
409  
410  
411  
412  
413  
414  
415  
416  
417  
418  
419  
420  
421  
422  
423  
424  
425  
426  
427  
428  
429  
430  
431  
432  
433  
434  
435  
436  
437  
438  
439  
440  
441  
442  
443  
444  
445  
446  
447  
448  
449  
450  
451  
452  
453  
454  
455  
456  
457  
458  
459  
460  
461  
462  
463  
464  
465  
466  
467  
468  
469  
470  
471  
472  
473  
474  
475  
476  
477  
478  
479  
480  
481  
482  
483  
484  
485  
486  
487  
488  
489  
490  
491  
492  
493  
494  
495  
496  
497  
498  
499  
500  
501  
502  
503  
504  
505  
506  
507  
508  
509  
510  
511  
512  
513  
514  
515  
516  
517  
518  
519  
520  
521  
522  
523  
524  
525  
526  
527  
528  
529  
530  
531  
532  
533  
534  
535  
536  
537  
538  
539  
540  
541  
542  
543  
544  
545  
546  
547  
548  
549  
550  
551  
552  
553  
554  
555  
556  
557  
558  
559  
560  
561  
562  
563  
564  
565  
566  
567  
568  
569  
570  
571  
572  
573  
574  
575  
576  
577  
578  
579  
580  
581  
582  
583  
584  
585  
586  
587  
588  
589  
590  
591  
592  
593  
594  
595  
596  
597  
598  
599  
600  
601  
602  
603  
604  
605  
606  
607  
608  
609  
610  
611  
612  
613  
614  
615  
616  
617  
618  
619  
620  
621  
622  
623  
624  
625  
626  
627  
628  
629  
630  
631  
632  
633  
634  
635  
636  
637  
638  
639  
640  
641  
642  
643  
644  
645  
646  
647  
648  
649  
650  
651  
652  
653  
654  
655  
656  
657  
658  
659  
660  
661  
662  
663  
664  
665  
666  
667  
668  
669  
670  
671  
672  
673  
674  
675  
676  
677  
678  
679  
680  
681  
682  
683  
684  
685  
686  
687  
688  
689  
690  
691  
692  
693  
694  
695  
696  
697  
698  
699  
700  
701  
702  
703  
704  
705  
706  
707  
708  
709  
710  
711  
712  
713  
714  
715  
716  
717  
718  
719  
720  
721  
722  
723  
724  
725  
726  
727  
728  
729  
730  
731  
732  
733  
734  
735  
736  
737  
738  
739  
740  
741  
742  
743  
744  
745  
746  
747  
748  
749  
750  
751  
752  
753  
754  
755  
756  
757  
758  
759  
760  
761  
762  
763  
764  
765  
766  
767  
768  
769  
770  
771  
772  
773  
774  
775  
776  
777  
778  
779  
780  
781  
782  
783  
784  
785  
786  
787  
788  
789  
790  
791  
792  
793  
794  
795  
796  
797  
798  
799  
800  
801  
802  
803  
804  
805  
806  
807  
808  
809  
810  
811  
812  
813  
814  
815  
816  
817  
818  
819  
820  
821  
822  
823  
824  
825  
826  
827  
828  
829  
830  
831  
832  
833  
834  
835  
836  
837  
838  
839  
840  
841  
842  
843  
844  
845  
846  
847  
848  
849  
850  
851  
852  
853  
854  
855  
856  
857  
858  
859  
860  
861  
862  
863  
864  
865  
866  
867  
868  
869  
870  
871  
872  
873  
874  
875  
876  
877  
878  
879  
880  
881  
882  
883  
884  
885  
886  
887  
888  
889  
890  
891  
892  
893  
894  
895  
896  
897  
898  
899  
900  
901  
902  
903  
904  
905  
906  
907  
908  
909  
910  
911  
912  
913

Db 162 nalagsanmrtiqvddvifngntyglrsrfsvgsnglgrsgmialggksdaftdtg-sig 220

```

Qy 221 VLYGHSRVSVAQNTYRVGGGQGHGICNFGAAYLERRKQRYFVQEGALKFNDSCKWERDLQR 280
D 221 vmaavsgsvysnfsngsg----- 239
Qy 281 QOWKYKPKYNNQELQYIEGDKSWRENLAPOYDTPIDPSLKKOOSAGNLFKLEY-D 339
D 240 -----inskef-----gydk-----ymkqnpksqlykmdirp 266
Qy 340 GFVFNKYAQFRDLNLTGKISRRIINRNQFNGLSLNPTNL---NLTAAYNSGKQYKPKG 396
D 267 defnsfelsartyenkftrditsddyyiky--hytfselidfnvtastergmqkyrdg 324
Qy 397 SKFTGWGLLKDFETNNAKILDLNNTATFRLPRETELQTLTGFNFYHNEYGKRNFPPELG 456
D 325 slytlyf---ktsagnsdsalidnntsrftv-adndlefmlgsklmtry----- 369
Qy 457 LFFDGPDDNGLYSLYGRFGDKGLLPQKSTIVOPASQVFNFTFYFAALKKKDI----YR 512
D 370 -----drtihsaagdpkanqesi--ennpfapsgqgdlsalytglkvtrgiweadfn 419
Qy 513 LNYSTNTV-GYRFGGEY-----TGYYSDDDFKRAFGENSPYKHKHCNRSKCGIYEPVLK 565
D 420 lnytrnritgypacdsvrvcvpggsydid- 450
Qy 566 KYGKKRANNSHVSISADFGDYFMPFASYSRTHRPNIQEMFYSGIDSGVHTALKPERAN 625
D 451 ---keggfnpsvqlsaqvtplqlpfygsksmrpnqemffnsgsgamnpflkperae 507
Qy 626 TWQGFNPTKYLKQDDTLGLKLVGRSRIDNIHN-----VYGRKWD 669
D 508 twqgfnldtrdliveqdalfkalayrsriqnyiseylvcsgrgrkcslpevigngwe 567
Qy 670 LNGDIPSWSTGLAYTIQHNRNFKDKVHKHGFLELNDYGRFPTNLISYAVOKSTOPTNF 729
D 568 ---gisdeysdmnyiyv---nsasdiakgfelemdydagfagrfsqgqddpctsi 620
Qy 730 SDASESPNNASKEDQLKGGYGLSRVSALPRDYGRLEVYGRWLNKLTGLGAMRYFGKSIR 789
D 621 asth-----fgagditelpkymtldtgvfrfdnaltlgtlikiytkarr 665
Qy 790 ATAERVIDGTNGONTNFRQLGRSISIKQETLARQLPIEDFYAAYEPKKNLIFRAEVKN 849
D 666 lspdeqentga-----iikq--dlpqiptliidlygtieynrnlitklsvqn 711
Qy 850 LFDRLYIDPLDAGNDAATQRYISSDFDKDKDEDVTCNADKTLGNGKYGGTSKSVLTNFA 909
D 712 lmrndysealnkn-----mmpglgdethpans-----ar 741
Qy 910 GRTEFLITMSYKF 921
D 742 grtwifggdirf 753

RESULT 14
AA07476
ID AA07476 standard; Protein; 915 AA.
xx
AC AA07476;
xx
XX 17-AUG-1999 (first entry)
DE
DE N.gonorrheae transferrin binding protein.
xx
KW Transferrin binding protein; iron; outer membrane protein; uptake;
KW Neisseria gonorrheae; nutrient; growth; Neisseria meningitidis; pathogen;
KW receptor; antibody.
xx
OS Neisseria gonorrheae.
xx
PN US5912336-A.
XX 15-JUN-1999.
PD

```

```

xx 23-DEC-1994; 94US-0363124.
xx
xx 23-DEC-1994; 94US-0363124.
PR 23-AUG-1990; 90US-0572187.
PR 03-NOV-1992; 92US-0973336.
PR 20-SEP-1993; 93US-0124254.
xx
PA (UUNC-) UNIV NORTH CAROLINA.
xx
PI Cornelissen CN, Sparling PF;
xx
DR WPI; 1999-357219/30.
DR N-PSDB; AAX78929.
xx
PT Nucleic acid molecules encoding transferrin binding proteins
xx
PS Disclosure; Fig 1; 38pp; English.
xx
CC This sequence represents the transferrin binding protein from Neisseria
CC gonorrheae. Transferrin is an iron-regulated, outer membrane protein
CC involved in uptake of iron, an essential nutrient for the growth of
CC N.gonorrheae and N.meningitidis. A similar sequence from N.meningitidis
CC is shown in AA07477. N.gonorrheae and N.meningitidis are two pathogens
CC of the genus Neisseria that are genetically similar, but pathologically
CC different. The growth of these cells can be inhibited by reducing the
CC ability of these cells to take up iron, e.g. by blocking the transferrin
CC receptor function. The transferrin binding proteins can be used to
CC raise antibodies for inhibiting the activity of the transferrin
CC receptor.
xx
xx Sequence 915 AA;
xx
Query Match 5.8%; Score 273.5; DB 20; Length 915;
Best Local Similarity 21.3%; Pred. No. 4.6e-11;
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;
Qy 5 FRLKPICFILMGVTLHYHYSAEDAGRAGSBAQIQVLEDVHVAK--RVPDKKVFETDARA 62
D 7 frnlclslmtal---payaenv-qag-qaqekqldtiqvkkakkqktrrdnevtglgkl 61
Qy 63 VSTRQDIFKSE-NLDNIVRSIFP-AFTQDQK--SSGIVSLNIRGDSGFGVRVMTVDGIT 118
D 62 vktadtlskeqvldirdltrydpgiavvegrgassg---ysirg-mdknrvslitvdgla 117
Qy 119 Q-TFYSTSTTAG--RAGSSQFGASVDSNFIAGLDVVVKGSFSGSAGINSIAGSNLRTLG 175
D 118 qlqsytaqaalgtrtagssgaineieyenvkaveiskgsnveqsggalagsvafqtk 177
Qy 176 VDDVQGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASGVLYGHSRVSVAQNYR 235
D 178 addvigeqrwgi-----qsktaysgknrgl-----tgsialagr 212
Qy 236 VGGG-----GQHIGNFGAAYLERRKQRYFVQEGALKFN-----SDSGKWERDLQRQW 283
D 213 iggaecallirtgrhageiraheaagr-----gvqsfurlapvdgdkgyayfiivee- 263
Qy 284 KYRKYNNQNEL---QKYIEGHDK---SWRENLAPOYDTPIDPSLKKOOS-----AG 331
D 264 ----ckngghekckanpkdkdvgedkrqvtvstrdytgnrfla--dplesyeswifrgp 317
Qy 332 NLF--KLEVDGVFNKYTAQ--FRDLNLTGKISRKI-----NRNQN--YG 371
D 318 frfenkrhyigllertqtdrtdmtvpafltkavfdanqkagslrgngkyagnhkyg 377
Qy 372 LSLNPTNLNLTAAAYNSGKQYKPKGKFTGWLLKQFETYNNAKILDLNNTATFRLPRET 431
D 378 -----glftsgennapvgaey-gtgvfyd-----et 402
Qy 432 ELQTTIGFNYFNEGYCKNRPPELGLFFD--GPDQNGL----- 458
D 403 htkrsrygleyvynadkdtwadyarlsydrqgigldhfhqthcsadgskycrpsadkp 462

```





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:54:36 ; Search time 42.33 Seconds  
(without alignments)  
2090.675 Million cell updates/sec

Title: US-09-762-926-6  
Perfect score: 4894  
Sequence: 1 MRSSFRLKPICFYLMGVTLX.....SVLTNFARGRTFLTMSYKF 921

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR\_71.\*  
1: pir1.\*  
2: pir2.\*  
3: pir3.\*  
4: pir4.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4883	99.8	921	2 E81076	TonB-dependent rec
2	4763.5	97.3	922	2 G81865	probable outer mem
3	2505	51.2	913	2 G64110	hypothetical prote
4	1096.5	22.4	755	2 A81436	probable outer mem
5	435.5	8.9	891	2 H83218	heme acquisition p
6	428.5	8.8	989	2 C83035	hypothetical prote
7	419.5	8.6	830	2 AH0477	probable TonB depe
8	359	7.3	852	2 C98310	hasR protein (Y089
9	359	7.3	923	2 AH2972	heme receptor hasR
10	309.5	6.3	784	2 B2437	TonB receptor-rela
11	273.5	5.6	744	2 B64049	outer membrane pro
12	273.5	5.6	766	2 D97634	probable outer mem
13	273.5	5.6	766	2 AF2857	outer membrane hem
14	273.5	5.6	915	2 A43335	transferrin-binding
15	269	5.5	908	2 JN0819	transferrin-binding
16	268.5	5.5	698	2 E82443	heme transport pro
17	265	5.4	791	2 F81056	hemoglobin recepto
18	265	5.4	910	2 C81832	transferrin-binding
19	256.5	5.2	911	2 JN0821	transferrin-binding
20	254.5	5.2	915	2 F81196	transferrin-binding
21	249	5.1	940	2 S49087	lactoferrin bindin
22	246	5.0	792	2 S61335	hemoglobin recepto
23	242.5	5.0	943	2 G81070	lactoferrin-binding
24	239.5	4.9	764	2 H83055	probable outer mem
25	239	4.9	944	2 C81798	lactoferrin bindin
26	238.5	4.9	912	2 C64107	transferrin-binding
27	233	4.8	723	2 C64058	outer membrane pro
28	230.5	4.7	912	2 S70501	transferrin-binding
29	228.5	4.7	877	2 AC2211	heme transport pro

30	225	4.6	725	2 A57148	outer membrane pro
31	225	4.6	911	2 S70511	transferrin-binding
32	222.5	4.5	953	2 B64083	hemoglobin-binding
33	221	4.5	843	2 A87275	TonB-dependent rec
34	219.5	4.5	914	2 S70906	transferrin-binding
35	210	4.3	931	2 S66574	transferrin-binding
36	208	4.3	810	2 A81965	hemoglobin-binding
37	206	4.2	1084	2 B64088	hemoglobin-binding
38	201	4.1	720	2 S58133	Fe-regulated prote
39	197	4.0	851	2 A83484	probable heme util
40	193	3.9	713	2 A56268	Fe-regulated prote
41	192.5	3.9	879	2 E71811	probable iron-regu
42	191	3.9	660	2 D91176	heme utilization/t
43	191	3.9	650	2 E86022	outer membrane hem
44	191	3.9	958	2 A82583	conserved hypochet
45	186	3.8	686	2 S28042	hemin receptor pre

ALIGNMENTS

RESULT 1

E81076  
TonB-dependent receptor NMB1497 [imported] - Neisseria meningitidis (strain MC58 sero  
C;Species: Neisseria meningitidis  
C;Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C;Accession: E81076  
R;Tettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B. ri, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Massignani, V.; Pizza, M. Science 287, 1809-1815, 2000  
A;Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; A;Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58. A;Reference number: A81000; MUID:20175755  
A;Accession: E81076  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-921 <FET>  
A;Cross-references: GB:AE002499; GB:AE002098; NID:g7226737; PIDN:AAF41853.1; PID:g722 A;Experimental source: serogroup B, strain MC58  
C;Genetics:  
A;Gene: NMB1497

Query Match 99.8%; Score 4883; DB 2; Length 921;  
Best Local Similarity 99.8%; Pred. No. 3.2e-298;  
Matches 919; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRSSFRLKPICFYLMGVTLXHYSAEDAGRAGSEAOIQVLEDVHVAKRVPKDKKVFETA 60

Db 1 MRSSFRLKPICFYLMGVTLXHYSAEDAGRAGSEAOIQVLEDVHVAKRVPKDKKVFETA 60

QY 61 RAVSTRODIFKSSNLNIVRSIPCAFTQDDKSGIYSLNIRGDSGFRVNTVMDGITQT 120

Db 61 RAVSTRODIFKSSNLNIVRSIPCAFTQDDKSGIYSLNIRGDSGFRVNTVMDGITQT 120

QY 121 FYSTTDAGRAGSSQFCASVDSNFIAGLDVYKGSFGSAGINSLAGSANLRTLGVDVV 180

Db 121 FYSTTDAGRAGSSQFCASVDSNFIAGLDVYKGSFGSAGINSLAGSANLRTLGVDVV 180

QY 181 QGNTYGLLLKGLTGTNSTKGNMAAIGARKWLESAGSVGLYGHRSRVAQNYRVGGG 240

Db 181 QGNTYGLLLKGLTGTNSTKGNMAAIGARKWLESAGSVGLYGHRSRVAQNYRVGGG 240

QY 241 QHIGNFCAEYLERKQRYFVQEGALKFNDSGKWERDLQROWKYKPKYNNQELQYI 300

Db 241 QHIGNFCAEYLERKQRYFVQEGALKFNDSGKWERDLQROWKYKPKYNNQELQYI 300

QY 301 EGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNFKIGSRK 360

Db 301 EGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNFKIGSRK 360

QY 361 IINRNYQFNGLSNPVTNLNTAAYNSGRQKPKGSKFTGWGLLKDPFYNNNAKILDNL 420

Db 361 IINRYQFNYGLSNPYTNLNLTAAYNSGRQYKPKSGFTGWGLLKDFETYNNAKILDN 420  
QY 421 NTATFRPLPRETELQTTLGFNYPHNEYKGNRPPEELGFFDGPDOODNGLYSLGRFKDGK 480  
Db 421 NTATFRPLPRETELQTTLGFNYPHNEYKGNRPPEELGFFDGPDOODNGLYSLGRFKDGK 480  
QY 481 LLPQKSTIVPAGSQYFNTFYDAALKKDIYRLNYSNTVYRGFGGYTGYGSDDEFKR 540  
Db 481 LLPQKSTIVPAGSQYFNTFYDAALKKDIYRLNYSNTVYRGFGGYTGYGSDDEFKR 540  
QY 541 AFGNSPTYKHKCNRSOGIYEPVLLKYGKKRANNSHSVISADFGDYFMPFASYSRTHRM 600  
Db 541 AFGNSPTYKHKCNRSOGIYEPVLLKYGKKRANNSHSVISADFGDYFMPFASYSRTHRM 600  
QY 601 NIQEMYTSOIGDSVGHVHTALPERANTWQFGFNTYKGLLKQDDPGLGLKLYGYSRIDNYI 660  
Db 601 NIQEMYTSOIGDSVGHVHTALPERANTWQFGFNTYKGLLKQDDPGLGLKLYGYSRIDNYI 660  
QY 661 HNVYCKWMDLNGDIPSWSSSTGLAYTIQHRNFKDKVHKHGFELNLDYDGRFETNLSYAY 720  
Db 661 HNVYCKWMDLNGDIPSWSSSTGLAYTIQHRNFKDKVHKHGFELNLDYDGRFETNLSYAY 720  
QY 721 QKSTQPTNFSADSPNNAKEDQKQGYGLSRVSALPRDYGRLEVTGRLGNKLTLLGGA 780  
Db 721 QKSTQPTNFSADSPNNAKEDQKQGYGLSRVSALPRDYGRLEVTGRLGNKLTLLGGA 780  
QY 781 MRYFGKSI RATAEERYIDGTNGNTSNFRQLGKRISIKQETTLARQPLIFDFYAAAYEPKN 840  
Db 781 MRYFGKSI RATAEERYIDGTNGNTSNFRQLGKRISIKQETTLARQPLIFDFYAAAYEPKN 840  
QY 841 LIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLGNGKYGTS 900  
Db 841 LIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLGNGKYGTS 900  
QY 901 KSVLTNFARGRTFLITMSYKF 921  
Db 901 KSVLTNFARGRTFLITMSYKF 921

RESULT 2  
G81865  
Probable outer membrane substrate binding protein NMA1700 [imported] - Neisseria meningitidis  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: G81865  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel  
; Holroyd, S.; Jagers, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,  
Nature 404, 502-506, 2000  
A:title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.  
A:Reference number: A81775; MUID:2022256  
A:Accession: G81865  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-922 <PAR>  
A:Cross-references: GB:AL162756; GB:AL157959; NID:g7380091; PIDN:CAB84928.1; PID:g738034  
A:Experimental source: serogroup A, strain 22491  
C:Genetics:  
A:Gene: NMA1700

Query Match 97.3%; Score 4763.5; DB 2; Length 922;  
Best Local Similarity 97.5%; Pred. No. 1e-290;  
Matches 899; Conservative 9; Mismatches 13; Indels 1; Gaps 1;

QY 1 MRSSFLRKPCFYLMGVTLTHYSAEDAGRAGSEAOQVLEDVHVAKRVPKDKKVFDTA 60  
Db 1 MRSSFLRKPCFYLMGVTLTHYSAEDAGRAGSEAOQVLEDVHVAKRVPKDKKVFDTA 60  
QY 61 RAVSTRODIFKSSNLDNIYRIPGAFQODKSSGIVSLNIRGDSGGRVNTWVDGTTOT 120  
Db 61 RAVSTRODIFKSSNLDNIYRIPGAFQODKSSGIVSLNIRGDSGGRVNTWVDGTTOT 120

QY 121 FYSTSDAGRAGSSQFSGASVDSNFIAGLDVWKGFSFGSAGINSLAGSANLRTIGVDVV 180  
Db 121 FYSTSDAGRAGSSQFSGASVDSNFIAGLDVWKGFSFGSAGINSLAGSANLRTIGVDVV 180  
QY 181 QGNNTYGLLLKGLTGTNTSTKGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGGG 240  
Db 181 QGNNTYGLLLKGLTGTNTSTKGNMAAIGARKWLESASGVLYGHSRRSVAQNYRVGGGG 240  
QY 241 QHIGNTGAELYERKKORYFVOEGALKFNSDGKWERDLQROOMKYKPKYKNYN-NOELQKY 299  
Db 241 QHIGNTGAELYERKKORYFVOEGALKFNSDGKWERDLQROOMKYKPKYKNYN-NOELQKY 300  
QY 300 IEHDKSWRENLAPOQYDITPIDPSSLKQKQAGNLFKLEYDGVFNKYTAQFRDLNKTIGSR 359  
Db 301 IEHDKSWRENLAPOQYDITPIDPSSLKQKQAGNLFKLEYDGVFNKYTAQFRDLNKTIGSR 360  
QY 360 KIINRYQFNYGLSNPYTNLNLTAAYNSGRQYKPKSGFTGWGLLKDFETYNNAKILD 419  
Db 361 KIINRYQFNYGLSNPYTNLNLTAAYNSGRQYKPKSGFTGWGLLKDFETYNNAKILD 420  
QY 420 NNTATFRPLPRETELQTTLGFNYPHNEYKGNRPPEELGFFDGPDOODNGLYSLGRFKDGK 479  
Db 421 NNTATFRPLPRETELQTTLGFNYPHNEYKGNRPPEELGFFDGPDOODNGLYSLGRFKDGK 480  
QY 480 GLLPQKSTIVPAGSQYFNTFYDAALKKDIYRLNYSNTVYRGFGGYTGYGSDDEFK 539  
Db 481 GLLPQKSTIVPAGSQYFNTFYDAALKKDIYRLNYSNTVYRGFGGYTGYGSDDEFK 540  
QY 540 RAGENSPTYKHKCNRSOGIYEPVLLKYGKKRANNSHSVISADFGDYFMPFASYSRTHRM 599  
Db 541 RAGENSPTYKHKCNRSOGIYEPVLLKYGKKRANNSHSVISADFGDYFMPFASYSRTHRM 600  
QY 600 PNQEMYTSOIGDSVGHVHTALPERANTWQFGFNTYKGLLKQDDPGLGLKLYGYSRIDNY 659  
Db 601 PNQEMYTSOIGDSVGHVHTALPERANTWQFGFNTYKGLLKQDDPGLGLKLYGYSRIDNY 660  
QY 660 IHNVYCKWMDLNGDIPSWSSSTGLAYTIQHRNFKDKVHKHGFELNLDYDGRFETNLSYA 719  
Db 661 IHNVYCKWMDLNGDIPSWSSSTGLAYTIQHRNFKDKVHKHGFELNLDYDGRFETNLSYA 720  
QY 720 YKSTQPTNFSADSPNNAKEDQKQGYGLSRVSALPRDYGRLEVTGRLGNKLTLLG 779  
Db 721 YKSTQPTNFSADSPNNAKEDQKQGYGLSRVSALPRDYGRLEVTGRLGNKLTLLG 780  
QY 780 AMRYFGKSI RATAEERYIDGTNGNTSNFRQLGKRISIKQETTLARQPLIFDFYAAAYEPKN 839  
Db 781 AMRYFGKSI RATAEERYIDGTNGNTSNFRQLGKRISIKQETTLARQPLIFDFYAAAYEPKN 840  
QY 840 NLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLGNGKYGTS 899  
Db 841 NLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEDVTCNADKTLGNGKYGTS 900  
QY 900 KSVLTNFARGRTFLITMSYKF 921  
Db 901 KSVLTNFARGRTFLITMSYKF 922

RESULT 3  
G64110  
Hypothetical protein H1217 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C:Accession: G64110  
R:Reischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.  
Science 269, 496-512, 1995  
C:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter  
A:title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: G64110  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA

A:Residues: 1-913 &lt;TIGR&gt;

A:Cross-references: GB:U32801; GB:L42023; NID:gl574143; PIDN:AAC22870.1; PID:gl574147; T

```
Query Match      51.2%; Score 2505; DB 2; Length 913;
Best Local Similarity 51.5%; Pred. No. 4.4e-149;
Matches 495; Conservative 150; Mismatches 228; Indels 88; Gaps 16;

QY 1 MRSSFRKLPICFYL---MGVTLXHYSAEDAGRAGSEAIOVLEDVHVKAARVPKDKKVF 57
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 1 MKKAIKNLITPLGINTIGMTITQQAETLQ-----QIDVVEKV-----ISNDKKPF 48

QY 58 TDARAVSTRODIFKSSNLDNVSIPCAFTQODKSSGIVSLNRGDSGRVNTWDGI 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 49 TEAKASTRENVFRETQIDQVIRSIIPCAFTQODKSGVSVNIRGNGLRVNTWDGV 108

QY 118 TOTFYSTADAGRAGSSQFGASVDSNFIAGLDVVKGSGSAGINSLAGSANLRTLGVD 177
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 109 TOTFYSTALDSGSGSSQFGAAIDPFIAGVDVYKNSFGASINALAGSANFRTLGVN 168

QY 178 DVVGNNYGLLLKGLTGNTSKGNAMAATGARKWLESAGSVGLYGHRSRVSQAQNYRVG 237
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 169 DVITDDKPPFIILKGMTSNATKSNFTMAAGRKWLDNGGVGVVYGSQREVSDYRI- 227

QY 238 GGGGHIGNFGAEYLERKQRYFVQEGALKFNSDSGKWERDLQROQWK-YPKYKNYN--- 293
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 228 GGGERLASLGODILAKEAYFRNAGYL--NPEGQWTPDLKSKHWSCKNPDYQKNGDCS 285

QY 294 -----QELQK---YIEGHDKSWRENLAPOYDITPIDP 322
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 286 YYRIGSAAKTPREILQELLNGKKPKDIEKLOKNGDIEETDKSFERN-KDQYSVAIEP 344

QY 323 SSLKQOQAGNLFKLEYDGVFNKYTAQFRLNLTGKSRKIINRNQYFNGLSLNPTYNLNL 382
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 345 GSLOQSRSHLLKPEYGDHQLCAQLRTLNDKIGSRKIENRNQYVFNENNSYLDLNL 404

QY 383 TAAVNSGRQKYPKSGKSTGWLKDFETYNNAKTILDNNNTATFPLPRETELOTTLGNYP 442
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 405 MAANIGKTYPKGGFAGQVADKLTKNVANTVDINNSTFLLPKIEIDLKTLTGNFY 464

QY 443 HNEYKGRFPEELGFFDPDQDNGLYS--LGRFKGDKGLLPKOKSTIVQAGSOYENTF 500
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 465 TNEYSKNRFPEELSLFYNDASHDGLYSHSKRGYSCTKSLLPORSVILQPSGKQKFTV 524

QY 501 YFDAALKKDIYRLNYSNTVGYRGGEYTYGYSDDEKRAFGNSPTYKHKHNRSCGIY 560
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 525 YFDTALSCKGIYHLNYSVNFTHYAFNGEYGVY-----ENTAGQQ-----IN 564

QY 561 EPVLKKGKGRANNSYSISADFGYEMPFPASYSTRHRMNIQEMYFSQJGDSGVHTALK 620
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 565 EPIUHKSHGKKAFFNHSATLSAELSDYEMPFPFTYSTRHRMNIQEMFFSOVSAGVNTALK 624

QY 621 PERANTWQGFNTYKGLLKQDDPLGLKLVYRSRIDNYIHNVYKWDNLGNDIPSWVSS 680
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 625 PEQSDTYQLGENTYKGLFTQDDVLGVKLVYRSFIFKNIHNVYGVWV--RDGNPTWAES 682

QY 681 TGLAYTTOHRNFKDKVHKHGFELNXYDGRFFTNLSYAKQSTQNFSDASESPNNAS 740
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 683 NGFYTTAHQNYKPIVKKSGVELEINYDMGRFFANVSAYQORTNOPTNYADASPRNNAS 742

QY 741 KEDOLKQGYGLSRVSALPRDYGRLVETRWLGNKLTGGAMRYFGKIRATAERYIDGT 800
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 743 QEDLLKQGYGLSRVMLPKDYGRLVETRWLGNKLTGGAMRYFGKIRATAERYIDGT 802

QY 801 NGGNTSFRQLGKRSIKQTETLARQPLIFDYAAEYKKNLIFRAEYKKNLFDRIYIDPLD 860
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 803 R-FKKNTLRRENYAVAKTEDIKKQPIILLDHSVSEYPIKLLIKAQVONLLDKRYVDPLD 861

QY 861 AGNDAATQRYSSDPDKDDEDVTCNPADKTLCLNGKYGGTSKSVLTNPFARGTFLITMSYK 920
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 862 AGNDAASQRYSSSL-----NNSTECAQDSSAC-----GGSDKTVLYNPFARGRTYILSLNYK 912

QY 921 F 921
```

Db 913 F 913

```
RESULT 4
A81436
Probable outer membrane siderophore receptor Cj0178 [imported] - Campylobacter jejuni
C;Species: Campylobacter jejuni
C;Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 31-Mar-2000
C;Accession: A81436
R;Parkhill, J.; Wren, B.W.; Mungall, K.; Ketley, J.M.; Churcher, C.; Basham, D.; Chil
C.W.; Quail, M.; Rajandream, M.A.; Rutherford, K.M.; VanVleet, A.; Whitehead, S.; Ba
Nature 403, 665-668, 2000
A;Title: The genome sequence of the food-borne pathogen Campylobacter jejuni reveals
A;Reference number: A81250; MUID:20150912
A;Accession: A81436
A;Status: preliminary
A;Molecule type: DNA
A;Residues: 1-755 <FAR>
A;Cross-references: GB:AL139074; GB:AL111168; NID:g967505; PIDN:CAB72661.1; PID:g696
A;Experimental source: serotype O2, strain NCTC 11168
C;Genetics:
A;Gene: Cj0178
```

```
Query Match      22.4%; Score 1096.5; DB 2; Length 755;
Best Local Similarity 31.1%; Pred. No. 7.3e-61;
Matches 288; Conservative 140; Mismatches 306; Indels 191; Gaps 25;
```

```
QY 12 FYLMGVTLHYHYSAEDAGRAGSEAIOVLEDVHVKAARVPKDKKVFDDARAVSTRODIFK 71
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 7 FCAVGLGCFASFASF-----ELEFDSLE-----ISGSKIKNDEKPFVTPGATSTREGIS 56

QY 72 SSENLDNVSIPCAFTQODKSSGIVSLNRGDSGRVNTWDGIQTFY-SYSTAGR 130
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 57 DTQSIDSVRSIPCTYNTDQAQGTQVQVNIKMSGFGVNTMIDGVQTFYFGSASDDPAR 116

QY 131 ---AGGSQFGASVDSNFIAGLDVVKGSGSAGINSLAGSANLRTLGVDVVGNNYTYG 187
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 FHSQGTGSFAFVADTNFLIGVDVTRGTGEGVGANALMGSAFNTKIGINDIVHDGNIFG 176

QY 188 LLKAGLTGNTSKGNAMAATGARKWLESAGSVGLYGHRSRVSQAQNYRVGSGQHIGNFG 247
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 177 FLGRFSGNSGIGFSPYSAVAGKTELENNYVGFALFUSAKRITQNTYTVGGG----- 229

QY 248 AEYLERRKQRYFVQEGALKFNSDSGKWERDLQROQWKYPKYKNYNQELQKYEIEGHDKSW 307
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 230 -----KIGSSWVDTDG-----KIGSSWVDTDG----- 240

QY 308 RENLAPQYDITPIDPSSLKQOQAGNLFKLEYDGVFNKYT---AOFRLNLTGKSRKIINR 364
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 241 -DGIA-DTNIAPFPDPLTQKPSQLFKLEY--VPNSFTNTIFSYRRYQNELAGRKIHND 296

QY 365 NYQFNYGLSNPYTNLNLTRAYNSGRQKYPKSGKSTGWLKDFETYNNAKTILDNNNTAT 424
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 297 NYQDWFHNPNWLNVLTAIYNOGIQTYGSKSTFAANDATANKAKNTATFTFDISDTLE 356

QY 425 FRLPRETELOTTLGFNFHNEYGNKRFPEELGFFDPDQDNGLYSYLGRFKGDKGLLPQ 484
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 357 GEWHR-FNLNIRFCANILLNDY-KNTLNTSI-----QGV--- 388

QY 485 KSTTVQAGSOYFNTFYDAAKKDIYRLNYSNTVGYRGGEYTYGYGSDDE-----FK 539
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 389 NSIPFQPRGKQNLFTYYLDNSLNYGIPTLDTNVNLLDWN---KGRHPACDEYFNCMCF 444

QY 540 RAFGENSPYKHKHNRSCGIYEPVLKKGKGRANNSHSVISADFGYEMPFPASYSTRHRM 599
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 445 KAAPD-----IDKGURL--NASVMSLAAAHIELITPVSFARTNRA 483

QY 600 PNIQEMYFSQJGDSGVHTALKPERANTWQGFNTYKGLKQDDTLGLKLVGYSRSDNY 659
   : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 484 PNQVEMFSSNEGINSFLKPEQANTWQGFNSFKHGLLKDDDRFGFKAYVYHTKIDY 543
```







QY 857 DPLDAGNDAATQRYSSDFDPKDKDEDTVCNADKTLONGKYGGTSKSVLTNFA--RGRTFL 914  
Db 806 -----VNYGDR-----LSFAPSGRGTIQ 823  
QY 915 ITMSYKF 921  
Db 824 GGFEYKF 830  
RESULT 8  
C98310  
hasR protein (Y08983) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)  
C:Species: Agrobacterium tumefaciens  
C:Date: 22-Oct-2001 #sequence\_revision 22-Oct-2001 #text\_change 11-Jan-2002  
C:Accession: C98310  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Ouellet, B.; Goldman, A.; Liu, F.; Wollam, C.; Allinger, M.; Dougherty, D.; Scott, C.; Lappas, C.; Markelz, B.; Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tumefaciens  
A:Reference number: A97359; PMID:11743194  
A:Accession: C98310  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-852 <KUR>  
A:Cross-references: GB:AE007870; PIDN:AAK90005.1; PID:g15159974; GSPDB:GN00170  
C:Genetics:  
A:Gene: AGR\_L 2875  
A:Map position: linear chromosome  
Query Match 7.3%; Score 359; DB 2; Length 852;  
Best Local Similarity 21.1%; Pred. No. 1.6e-14;  
Matches 200; Conservative 117; Mismatches 336; Indels 296; Gaps 37;  
QY 62 AVSTRQDIKSS--ENLDNIVRSIPGAFTQODKSS--GIVSLNIRGDSGFGFRVNTWVGITQ 119  
Db 111 SVVSREAIQAGVRNTRDFNVRVSGYAGEGNGSFPTSPNVRGLQESGRVVSIDGARQ 170  
QY 120 TF---YSTSTDRAGAGSSQFGAS-----VDSNFIAGLDVVKGSFSGSAGINSLAGSANL 171  
Db 171 NAQRGYST-----GGASIYSANNGQAYVDAAFIRAVEVEKMTSATSGNAGSLGKVEF 223  
QY 172 RTLGVDVVQGNNTYGLLLKGLTGTN--STKGNAIAAGARKWLESASGVVLYGHSRRS 229  
Db 224 RTVSAADLIPGANKGAENVNVRSGNGYDFQGSVLAIV-----REPDPGLSFVAGYS-RT 277  
QY 230 VAQNYRVGGGGQHI-----GNFG---AEYLERKQRYFVQEGA 264  
Db 278 IMDEYKIGTKGKALSTALTMKDLLGRDGSWSTFFKSGDFGVDVQTSLSWMHQNDVFOGAS 337  
QY 265 LKFNDSGK-----WERDLQROQWKYPKYKNYN--NOELQKYIEGHDKSWRENLAPO 314  
Db 338 TVVDREGVRNDSIVAKLMDWDPSELIDFKSSLWLDNMWTHLR-----AART 384  
QY 315 YDI-TPID-----PSSLKQOASAGNLFKLEYDGVFNKYTAQ--FRDLNTRKIGSKRIINRY 366  
Db 385 YAVETNLDMLGRSFGGSLNTR--FDTRAGALSILNYGAEAPRDIAATSVATSAIQNP 441  
QY 367 QF--NYGLSLNPYTNLNTAAYNSGRQPKYKSKFTG-----WGLLKDPETYNNAKILDL 419  
Db 442 SFASSY-TSFPAGRRDVSALFLNGELEPADWTTLGGVRYDMSRLKGSATYSYFK-----496  
QY 420 NNTATRLPRETELQTLGFTNYPHNRYGKNRPFPEELGFLFDGPDQDNGLYSLGKREKGD 479  
Db 497 -----ESIVTTSVPCDLVRNH-----513  
QY 480 GLLPQKSTIVQAGSYQNTFTYDAALKKDIYRLNYSNTNTVYRFGGYTGYGSDDEFK 539  
Db 514 -----TALEYFNQVFLPANLPVWASRYNVFLASIW-----543  
QY 540 RAFGENSPYKHKCNCSGIEYPLVKKYKGRANNHSVSIADFG-----DY 586  
Db 544 -----PRTSANCMPGTGI-----TTKTPVTEYPSHEVDIDRTYSAWLPSATIEFKPVDW 592

QY 587 FMFPASYSRTHRPNIQEMYF--SQIGDS-----GVHTALKPERANTWOFNFNTYKKGLL 639  
Db 593 FRYPVYSQSRLPPTILETFFAGARPCDSAGYEYAPNOSLRAEKATYIEIGANKSPDGLV 652  
QY 640 KQDDTLGLKLVGYSRIDNVI-----HNVYCKWDLNGIDIPSWSSSTGLATVIOHRNF 692  
Db 653 LDDDTLIRIKMAAFREVVDYIALGYLVTQVDFRTY-----TSFVNLGDGTTY-----699  
QY 693 KDKVHKHGFELNLDYDGRFETNLSYAYOKSTOPTNFSDASESPNNAKEDQLKQGYGLS 752  
Db 700 -----MRGELEGENYDARSFWIGGSATVLKTEWPEKTQVFSNSTTTTSGE-----744  
QY 753 RVSALPRDYGRLEVGTRWLGKLTGLGAMRYFGKISRATAEERYIDGTNGGNTSNFRQLG 812  
Db 745 -IVAWPGD-----VAPKM--KLTLDGGMRFDEKESLGARLNVHTPTQSRTLDTGCLN- 794  
QY 813 KRSIKOTELAROPILDFDYAAVEPKNLIIFRAEVKNLFDORRYIDPLDAGNDAATQRYYS 872  
Db 795 -REITDPYTTV-----DLYGSYAFNDKATLREAVNNLTDRKYIPAASA-----YT 838  
QY 873 SFDPKDKDEDTVCNADKTLGCKYGGTSKSVLTNFAARGRTFLITMSYKF 921  
Db 839 -----APGRTFIATMNVKF 852  
RESULT 9  
AH2972  
heme receptor hasR [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AH2972  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo-  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutyavin, T.; Levy, R.; Li, M.; McCl  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
ster, E.W.  
A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AH2972  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-923 <KUR>  
A:Cross-references: GB:AE008689; PIDN:AAL44198.1; PID:g17741777; GSPDB:GN00187  
A:Experimental source: strain C58 (Dupont)  
C:Genetics:  
A:Gene: hasR  
A:Map position: linear chromosome  
Query Match 7.3%; Score 359; DB 2; Length 923;  
Best Local Similarity 21.1%; Pred. No. 1.8e-14;  
Matches 200; Conservative 117; Mismatches 336; Indels 296; Gaps 37;  
QY 62 AVSTRQDIKSS--ENLDNIVRSIPGAFTQODKSS--GIVSLNIRGDSGFGFRVNTWVGITQ 119  
Db 182 SVVSREAIQAGVRNTRDFNVRVSGYAGEGNGSFPTSPNVRGLQESGRVVSIDGARQ 241  
QY 120 TF---YSTSTDRAGAGSSQFGAS-----VDSNFIAGLDVVKGSFSGSAGINSLAGSANL 171  
Db 242 NAQRGYST-----GGASIYSANNGQAYVDAAFIRAVEVEKMTSATSGNAGSLGKVEF 294  
QY 172 RTLGVDVVQGNNTYGLLLKGLTGTN--STKGNAIAAGARKWLESASGVVLYGHSRRS 229  
Db 295 RTVSAADLIPGANKGAENVNVRSGNGYDFQGSVLAIV-----REPDPGLSFVAGYS-RT 348  
QY 230 VAQNYRVGGGGQHI-----GNFG---AEYLERKQRYFVQEGA 264  
Db 349 IMDEYKIGTKGKALSTALTMKDLLGRDGSWSTFFKSGDFGVDVQTSLSWMHQNDVFOGAS 408  
QY 265 LKFNDSGK-----WERDLQROQWKYPKYKNYN--NOELQKYIEGHDKSWRENLAPO 314

Db 409 TVVDREGVRNDIVAKLDWDPESELIDFKSSLWLNNDNWTHEL- - - - -AART 455  
QY 315 YDI-TPID- - - - -PSSLKQOAGNLFKLEYDGVFNKYTAQ- FRDLNTKIGRKKIIRNY 366  
Db 456 YAVETNLDMLGRSFGSLENTSR- - - - -FDTRAGALSILNYGAEAFRIATSVATSATIAQNP 512  
QY 367 QF- - - - -NYGLSLNPTNLNTAAVNSGRQKYPKSKFTG- - - - -WGLKDFETNNAKILD 419  
Db 513 SFASSY-TSFSAGRRDVASLFLNGELEPADWITLGGVRYDMSRLKGSATYSEK- - - - -567  
QY 420 NNTATPRLPRETELQTLGFNFHNEYGNRPPEELGLFFDGDQDNGLYSYLGRFGDK 479  
Db 568 - - - - -ESIVTTSVPCDLVRNHY- - - - -584  
QY 480 GLLPQKSTIVQAGSYQYNTFTFYDAALKDKDIYRLNYSNTVYGRFGGYTGYGSDDEK 539  
Db 585 - - - - -TALEYFNQVLEPANLPVWASRYNVFLASIW- - - - -614  
QY 540 RAFGENSPYKKHCRSCGIYEPVLKKYKGRANNHVSISADFG- - - - -DY 586  
Db 615 - - - - -PRTSANCMPGTGI- - - - -TTKTPVTEYPSHEVIDIRTYSAWLPSTATIEFKPVDW 663  
QY 587 FMPFASYSRTHMPNIOEYF- - - - -SQIGDS- - - - -GVHTALKPERANTWQGFNTYKKGILL 639  
Db 664 FRPYYSQSLSRPPTEILETFAGARPGDSAGYEPNQSLRAEKATYIELGANMSFDGVL 723  
QY 640 KODDTLGLKLVYRSRIDNY- - - - -HNYYGKWDNLNGDIPSWSSVSTGLAYTIOHRNF 692  
Db 724 LDDDTLRKIAAFRRVKDYIALGYLVTDQVFDRTY- - - - -TSFVNLDGTTY- - - - -770  
QY 693 KDKVHKHGFELNVDYGRFFTNLSYAYOKSTQPTNFSASESPNNASKEDQLKQYGLS 752  
Db 771 - - - - -MRGLESGNDYDARFVWGGSVATVUKTEWPEKTQVFSSTTTSGE- - - - -815  
QY 753 RYSALPRDYGRLEVGTRWLGKNTLGGAMRYFCGKSTRATAERYIDGTNGGNTSNFRQLG 812  
Db 816 -IVAWPGD- - - - -VAPKM- - - - -KLITDGGMRFFDEKFSGLARLNHVHTPTQSRITLDTGNL- 865  
QY 813 KRSIKOTETLARQPLIFDYAYAEKPKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYYS 872  
Db 866 -REITDPYTTV- - - - -DLYGSYAFNDKATLRFVAVNNLTDKRYIPAAASA- - - - -YT 909  
QY 873 SFDPKDKEDVTCNADKTLGCKGYGTSKSVLTNFAARGRTFLITMSYKF 921  
Db 910 - - - - -APGRFTIATMNVKF 923

RESULT 10  
TonB receptor-related protein VCA0625 [imported] - Vibrio cholerae (strain N16961 serogr  
C:Species: Vibrio cholerae  
C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001  
C:Accession: D82437  
R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;  
chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, P  
l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.  
Nature 406, 477-483, 2000  
A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.  
A:Reference number: A82035; MUID:20406833  
A:Accession: D82437  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-784 <HEI>  
A:Cross-references: GB:AE004392; GB:AE003853; NID:g9658031; PIDN:AAF96526.1; GSPDB:GN001  
C:Genetics:  
A:Gene: VCA0625  
A:Map position: 2

Query Match 6.3%; Score 309.5; DB 2; Length 784;  
Best Local Similarity 21.2%; Pred. No. 1.8e-11;  
Matches 195; Conservative 122; Mismatches 343; Indels 261; Gaps 43;

QY 6 RLKPICFYLMGVTLHYHYSAEDAGRAGSEAIQVLEDVHVYKAKRVPKDKVFTDARAVST 65  
Db 36 QLKPL-FTLLPVVLSVVOAQE- - - - -NTEQAVDETIVH- - - - -GQSILTDQR- - - - -T 79  
QY 66 RQDIFKSSNLDNIVRSIPGAFQODKSSGIVS- - - - -LNTGRSGSGRVN 111  
Db 80 RSLDK- - - - -VRGIANA- - - - -DIFSGITSVQNNHNEAGALDGRVGOGEGRVP 127  
QY 112 TWVDGTQTFYSTDAGRAGSGSFGASVDSNFIAGLDVVKGS- - - - -FSGSAGINS 164  
Db 128 IFDGLSQ- - - - -SHTSRGYQGVSDR- - - - -TYIDTLLSLTVNKGATIESSPYASGAVG- - - - -179  
QY 165 LAGSANLRLQVDDVVQGNNTYGLLKG- - - - -LTGTNSTKGN-AMAAIGARKWL 213  
Db 180 -GVVNATTGIGIKDIKDQAFGVWLKARANHNRTPDVSGDYSEOGQVALDERGEHSF 237  
QY 214 ESGA- - - - -SVGLYGHSSRSVAQNYR- - - - -VGGGGOHIGNFGAEYLE 252  
Db 238 KHGSLMLGLGYOAESENTVLAYSKRSGNHFAKKGYEYOEYQEPVVGQGVVWTSFE- - - - -294  
QY 253 RRQRYFVOEGALKFNDSGSKWERDLQROQWKYKPKYKNYNNQELQYIEGHDKSWRENLA 312  
Db 295 - - - - -SDSWLFKLASDTG- - - - -TAHNADFNVRHHAQKAGEVLMAYWYKSSDWEGNPY 343  
QY 313 PQYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNTKIGSKRIINRYQFNGL 372  
Db 344 PD- - - - -GKDRMPQWGLGT- - - - -AKVNTYSA- - - - -NYTYQPDH- - - - -373  
QY 373 SLNPTYNLNLTAAYNSGRQKYPKSGFTG-WGLLKDFETYNNAKILDLNNTATFRLPRET 431  
Db 374 - - - - -PWLNLANFWYTTADL- - - - -AQYNGLMALGTNAEQYFHAY- - - - -HNDRSLSLTNET 422  
QY 432 EL-QTTLGFNYHNEYGNRPPEELGLFFDGDQDNGLYSYLGRFGDKGLLPKSTIVQ 490  
Db 423 LLTQWVRLNYGLAQNERLSPEE- - - - -DGQT- - - - -RF- - - - -TKTITS 458  
QY 491 PAGSQYNTFYFDAALKDKDIYRLNYSNTVYGRFGGYTGYGSDDEKFAFGENSESTYK 550  
Db 459 RHGKRTAQNLFANADIDYSPRLVOLGLNLHNAK- - - - -STDYQTKOOL- - - - -DYK 504  
QY 551 KHCNRCGIVEPVLLKYGKGRANNHVSISADFGDYFMP- - - - -FASYSRTHMPNIOEY 606  
Db 505 E- - - - -KLDLSEFTYALTPTQFLKSSRYRPPSLYETT 540  
QY 607 FS-QIGDSGVHTALKPERANTWQGFNTYKKGILLKODDTLGLKLVGYRSRIDNYIHNVY 665  
Db 541 LSNEVFSYPNPYKPEQAWNNEVGQVFMASNSVLQDDRLNLSVSYFRNSIKDFISG- - - - -G 598  
QY 666 KWDLLNGDIPSWSSVSTGLAYTIOHRNPKDKVHKHGFELNLYDYGRFFTNLSYAYOKSTQ 725  
Db 599 RLAKTPG-MSEW-QANFTFT- - - - -NY-DKLQLSGWELGAHYQYAWLYTHFAATLYSETK 650  
QY 726 PTNFSASESPNNASKEDQLKQCYGLSRVSALPRDYGRLEVGTRWLGKNTLGLGAMRYF- 784  
Db 651 ICSVQQAQVAESDTC- - - - -NSLGFAMWLTPTTRIPPKQNLNLYNVGKFFNDTLDSGVKVSYS 708  
QY 785 GKSIRATAERYIDGTNGGNTSNFRQLGKRSIKQETLARQPLI- - - - -FDYFAAYEP 837  
Db 709 GKS- - - - -NPSDWLAGT- - - - -AANPILEIPSDYIDLYSOYEL 742  
QY 838 KKNLIFRAEVKNLFDRIYIDP 858  
Db 743 NANTQFFFAINNVTDRYQVRP 763

RESULT 11  
B64049  
outer membrane protein hxcC homolog - Haemophilus influenzae (strain Rd Kw20)  
C:Species: Haemophilus influenzae  
C>Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C:Accession: B64049  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage

; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman, D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M. Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter, A.; Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: B64049  
A:Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA  
A:Residues: 1-744 <TIGR>  
A:Cross-references: GB:U32696; GB:U42023; NID:q1573057; PIDN:AAC21789.1; PID:q1573065; T

Query Match 5.6%; Score 273.5; DB 2; Length 744;  
Best Local Similarity 20.9%; Pred. NO. 3e-09;  
Matches 181; Conservative 136; Mismatches 366; Indels 183; Gaps 41;

**Qy** 14 LMGVTLHYSYAEDAGRAGSEAQIQVLEDVHVYKAKRPDKKVFTDARAVSTRQDIKF-S 72  
| : | : : : : | : : : : : : : : | : | : : : :

**Dd** 15 LVGIQL-NVTAKQNSSNSNREELPLIVTNTDDSNKLP-----GRSVLKQKNIEQXQ 65

Qy	73	SENLDNIVRSIPG	AFTQDKSSGI	VSINIRGDSG	FGRVNT	MDGITQT	FYSTST	DAGRAG	132
		: :   : :   :	:	:	:	:	:		
Db	66	ADNAANLINILP	GVNMAGGFR	PGGQTLN	INGMDAED	VRVQLD	GAATKSEK	-----	116

QY 133 GSSQFGASVDSNFIAGLDVVKGFSFGSAGINSLAGSANRLTGLGVDDVVGQNNITYCLLLKG 192  
 117 -YQGSIFTEPELLRRVTVDKGNYSQYNGGFGAGTVKFTKDAKDFLOENOKIGGFLLK- 174  
 Db

Qy	193	LTG	TNS	----	TKGN	AA	TG	ARK	W	LE	S	G	V	G	V	L	Y	G	H	S	R	R	S	S	V	A	Q	N	R	V	Y	V	G	G	G	G	H	I	G	N	F	G	247			
										:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:	:					
Db	175	-YGN	SN	NN	NN	N	O	K	T	S	T	A	L	V	L	O	N	E	O	K	N	I	D	L	----	LL	F	S	V	R	N	----	AG	D	Y	K	R	P	D	N	S	K	I	L	----	222

<b>Qy</b>	248 AEYLERRQRIFYOEALCKFNSSDSGKWERDLQRQQWYKKYPKNYNNOELQKYIEGHDKSW	307
<b>Db</b>	223 ---FSKNOK---TGLIKLN-----WOISP-----EHLATISSVVGTHKGW	257

<b>Qy</b>	308	RENLAPQYDITPIDPSSLKQOSAGNLFKLEVDGVFNKYTAQFRDLNTKGSRKIINRNYQ	367
		-   -       :   :   :   :   :   :   :   :   :   :   :   :   :   :   :   :	
<b>Db</b>	258	-EPEAAKRILP-----KPSLSDMRYGTD-IAWERKLGYRD-----OKDENYTLLK	301

```

368 FNYGLSLNPYTNLNLTAAYNSGRQK--YPKCSKFTGHGLL--KDFETYNNAKILDLNNTA 423
      :|| ||| ||| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :|| :||
302 YNYLPENNPWNLSTQFSYKTTQNDMPKPSAGSLVGSLSGNOSWITTYSDLTETDINNTS 360

```

```
QY      424 TFLPRETELQTTFGNYFNEYGNRPPEELGLFFDPDQDNGLYSYLGRFGDKGLLP   483
          || : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      361 TENI-KTIVHELLEGLWLKNT--RNT-----LMYDKSKVRKADNV-GYFO-----   403
```

**Qy** 484 QKSTIVPAGSQYNTFYFDAALK-KDIYRLNYSNTVGYRFGGEYTGYGSDDEFKRAF 542  
| | | | | : | | : | |  
**Db** 404 ---PYMPSGROTOAFYLQDOIKNKI---FST---GVRY-----DHINNIG 443

**Qy** 543 GENSPTYKKHCNRSCGIYEVLKKGKKRANH--VSISADFGDYFMPFAFSYSTHRMP 600  
:  
| : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :  
**Dh** 444 QKNIAL--KYNDI SAG-----HDYSOKNYNCWSVVICI NVYNVHYI STETNDCSTWDAE 495

Qy 601 NIQEMFYSGIDSGVHTA---LKPERANTWQFGENTYKGLLKQDDTLGLKLVGYRSRID 657

Dh 496 VTNEQVTEAFKQCSVDATSI NI EKWTNOMDVCC I TMI NUI EAENDA BQDDMVVDVNDQV 555

Qy 658 NYIH-----NVYGKWDLNGDI-PSWVSSTGLAYTIQHRNFQDKVHKHGFLELNYDYGR 711

QY 712 FFTNLSYAYQKSTQPTNFS DASESPNNASKEDQLKQGYLSRVSALPRDYGRLVGT RWL 771

QY 772 GNKLTLGGAMRYFGKSIRATAEERYIDGTNGCNTSNFRQLGKRISKQTETLARQPLIFDF 831

QY	832	YAAYPEK--KNLIFRAEVKNLFDRRY	855
		: :     : :	
Dd	696	FLSWSPAKIKGMNVKITVDNLFNRAY	721

RESULT 12

D97634  
probable outer membrane hemin receptor (PA4710) [imported] - Agrobacterium tumefaciens  
C:Species: Agrobacterium tumefaciens  
C:Date: 30-Sep-2001 #sequence\_revision 30-Sep-2001 #text\_change 11-Jan-2002  
C:Accession: D97634  
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurollo, B.; Goldm  
A: Liu, F.; Willam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz,  
Science 294, 2323-2328, 2001  
A:Title: Genome Sequence of the plant Pathogen and Biotechnology Agent Agrobacterium  
A:Reference number: A97359; PMID:11743194

A;Accession: D97634  
A;Status: preliminary  
A;Molecule type: DNA  
A;Residues: 1-766 <KUR>

A:Residues: 1-700 (AA)  
A:Cross-references: GB:AE007869; PIDN:AAK88029.1; PID:gl5157447; GSPDB:GN00169  
C:Genetics:  
A:Gene: AGR\_C\_4156  
A:Map position: circular chromosome

Query Match 5.6%; Score 273.5; DB 2; Length 766;  
Best Local Similarity 22.1%; Pred. No. 3.1e-09;  
Matches 219; Conservative 111; Mismatches 336; Indels 325

QY 13 YLMGVTLYHSY-----AEDAGRAGSEAQIVLEDVHVVKAKRVPKDKV 56  
 ||| :: } : ||| | | ||| |  
 Db 19 YLMSIRTHSALLCTAISLPIAGPARODA--ASOENGTTLEKTVVKGKRY-KSANA 75

QY 57 FTDA RAVS--TRQDIEK-----SSENLDNIVRSIPGAFTQQDKSSGIVSNIRGDSGFGRV 110

DP 76 AADTPLASOTTAEDIRKKDKGISIKDLGNTE--PGVDYVDSKPGPPGFIIRGLGG-ARV 132

QY 111 NTMWGIGITQTFYSTDTAGRAGGS-SQFGASVDSNFITAGLDVVVKGSFSGSAGINSLAGSA 169  
133 VTLLDNIPVYFNNEAROGAATTTLTSDTSSSEDFSSUSSDVVYRGADSSRIGSGALGGAL 192

QY 170 NLRTLGVDDVVQGNTTYGLLKGLTGNSTKNMAAATGARKWL----- 215  
 ||||| :||: :| | | : :| :  
 DP 193 VLRTLEPEDLIGEGKDVGWYAK--TSYDSEDSRVGGSLAAKKTIENSVLFQGSYKRNGE 250

QY 216 --GASVGVLYG--HSRRSVAQNYRVG-----GGQHGIGNFGAEYLERRKQRYFVQE 262

DP 251 TDNRGTADLYGTRRTKPNPADTYESNLFKIRDLEGCHRGILTAERYSLNRSDMKTLQ 310

QY 263 GALKFNSDGGKWERDLQRQWKYKPKYKNYNNQELQKYIEGHDKSWRENAPQYDITPIDP 322

Dh 311 GV-----SVSG-----SWRIGDVBGYEDEF-----RDRYSIDYEFVA--P 344

QY 323 SSLKQQSAGNLFKLEYDGVFNKYTAQFRDLNTKGSRKIINRNVQFNYGLSLNPYNLNL 382  
| : | | | : : : | : | : | :  
Db 345 SPSPCVIDRANI - SLVGMPI QWSSCAACDPIMNNCI V I D R E M B N C A F -----GT 391

```
Qy      383 TAAVNSGRQKYPKGS-----KFTGWGLLKDFETYNNAKILDNNNTATFRLPRETELQTTL 437
          ||| : : | : | : ||| : : :
Dk      202 VGGT ESG --- EET ACACGVMYDPFGCCNNAUMDRPGCEI PA-----NTAGCYMSSCGSCGDMDPVS 442
```

QY 438 GFN---YFNE--YGKNRFPPELGFFDGDQDNGLSYLGRFKGDKLLPQKSTIVQPA 492  
| : | : | : || | | | |  
442 GKEGLVVDPTATG-NCPRTDGTTCDCGVAV- - - - - DD 476

[illegible]

QY 553 CNRSCGIYEPVLKXYGKKRANHHSVISADFCGYFMPFASYSRTHMPNIQEMY--FSQI 610





```
Db 464 KSDRMIEESRNLQAVFKKAFDTAKIRHNLINLGYDRFKSQLSHSDYILQNAVQAYDL 523
Qy 507 -----KKDIYRLNYSNTVVG-----YRFGGE-YT-----GYGSDDEFFK 539
Db 524 ITPKKPPPPNGSKDNPYRVSVIGKTTVNTSPICRFGNNTYTDCTPRNIGGNGYAAVQDNV 583
Qy 540 RAFGENSPTYKKHCNRSCGIYEPVLKKYKKRANHSHVS-----ISADFGDIYMPFASY 593
Db 584 R-----LGRWADVAGIRYDIRSTHSEDKSVSTGTHRNLSWNAAGVVLKPFWTM 631
Qy 594 SRTH-----RMPNIQEMIFSOIGDSGVHTALKPERANTWQFG-----FNT 633
Db 632 DLTYRSTGFLPSPFAEMYGWRAGESLKTLDLPEKSFNREAGIVFKGDFGNLEASYFNN 691
Qy 634 YKGLLKODDTLGLKLVGYRSRIDN-----YIH-----NVYK--WBDLNG 672
Db 692 AYRDLI-----AFGYETRTQNGQTSASGDPGYRNAQNARIAGINILGKIDMHGVMG 742
Qy 673 DIPSWSVSTGLAY-----TIQRNFKDKVKHKGFELELNNDY--GRFTNL 716
Db 743 GLPDGLYST-LAYNRIRKVKDADIRADRTFVTSYLFDAVQPSRYVLGLGYDHPDGIWGINT 801
Qy 717 SYAYOKSTQPTNFSASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWLGKNKLT 776
Db 802 MFTYSK-----AKSVDEL-----LGSQAL 820
Qy 777 LGGAMRYFGKSIRATABERYIDGTNGGNTSNFROLGKRSIKQTETLARQPLIFDFYAAYE 836
Db 821 LNG-----NANAKKAASRR-----TRPWYVTVTDVSGYN 848
Qy 837 PKKNLIFRAEVKNLFDRIYI 856
Db 849 IKHLLTRAGVYNLLNRYV 868
```

Search completed: July 24, 2002, 08:54:48  
Job time: 539 sec





GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen-Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 09:00:10 ; Search time 20.76 Seconds  
(without alignments)  
1717.760 Million cell updates/sec

Title: US-09-762-926-6

Perfect score: 4894

Sequence: 1 MRSSRLKPCIFYLMGVTLX.....SVLTNFARGRTFLITMSYKF 921

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 105224 seqs, 38719550 residues

Total number of hits satisfying chosen parameters: 105224

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt\_40.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	2505	51.2	913	1 YC17_HAEIN	P45114 haemophilus
2	273.5	5.6	744	1 HXCI_HAEIN	P44523 haemophilus
3	273.5	5.6	915	1 TBPI_NEIGO	Q01996 neisseria g
4	269	5.5	908	1 TB12_NEIMB	Q06987 neisseria m
5	256.5	5.2	911	1 TB11_NEIMB	Q09056 neisseria m
6	249	5.1	1013	1 HGBA_HAEIN	Q9Kiv2 haemophilus
7	243.5	5.0	999	1 HGPE_HAEIN	O87296 haemophilus
8	242.5	5.0	943	1 LBPA_NEIMB	O67379 neisseria m
9	239.5	4.9	993	1 HGBC_HAEIN	Q9Kiv0 haemophilus
10	239.5	4.9	1063	1 HGPI_HAEIN	P44795 haemophilus
11	239	4.9	944	1 LBPA_NEIMA	Q9Jtk4 neisseria m
12	238.5	4.9	912	1 TBPI_HAEIN	P44970 haemophilus
13	233	4.8	723	1 Y262_HAEIN	P44600 haemophilus
14	228.5	4.7	1077	1 HGPA_HAEIN	Q9za21 haemophilus
15	227	4.6	999	1 HGP2_HAEIN	P44809 haemophilus
16	225	4.6	725	1 HXC2_HAEIN	P45357 haemophilus
17	209	4.3	810	1 HPUB_NEIMC	P96949 neisseria m
18	208	4.3	810	1 HPUB_NEIMA	Q9Jwa2 neisseria m
19	206	4.2	1084	1 HGPF_HAEIN	P44836 haemophilus
20	204.5	4.2	1067	1 HGBB_HAEIN	Q9Kiv1 haemophilus
21	201.5	4.1	1066	1 HGPC_HAEIN	Q9x442 haemophilus
22	196	4.0	1046	1 HHUA_HAEIN	Q48153 haemophilus
23	182.5	3.7	999	1 HGPA_HAEIN	Q57408 haemophilus
24	181	3.7	972	1 HGBB_HAEIU	Q47957 haemophilus
25	180	3.7	774	1 FECA_ECOLI	P13036 escherichia
26	179.5	3.7	676	1 HMUR_YERPE	Q56989 versinia pe
27	178.5	3.6	614	1 BTUB_ECOLI	P06129 escherichia
28	178.5	3.6	697	1 HEMR_YEREN	P31499 versinia en
29	176.5	3.6	1567	1 ICEN_XANCT	P18127 xanthomonas
30	175	3.6	700	1 YNCD_ECOLI	P76115 escherichia
31	173	3.5	972	1 HGBA_HAEIU	Q47952 haemophilus
32	172.5	3.5	710	1 FOXA_YEREN	Q01674 versinia en
33	165.5	3.4	614	1 BTUB_SALTY	P37409 salmonella

34 163 3.3 1104 1 COLA\_CLOPE P43153 clostridium  
35 162 3.3 1258 1 ICEN\_ERWHE P16239 erwinia her  
36 157 3.2 1322 1 ICEA\_PANAN P20469 pantoea ana  
37 156 3.2 815 1 FPVA\_PSEAE P48632 pseudomonas  
38 155.5 3.2 729 1 FHUE\_ECOLI P16869 escherichia  
39 153 3.1 809 1 PUPB\_PSEPU P38047 pseudomonas  
40 151.5 3.1 687 1 VIUA\_VIBCH Q00964 vibrio chol  
41 146.5 3.0 1148 1 ICEK\_PSEEX Q00611 pseudomonas  
42 145.5 3.0 1113 1 N116\_YEAST Q02630 saccharomyc  
43 144.5 3.0 989 1 SERA\_PLAFG P13823 plasmodium  
44 142 2.9 730 1 DCOR\_LACS3 P43099 lactobacill  
45 142 2.9 1034 1 ICEN\_PANAN Q47879 pantoea ana

#### ALIGNMENTS

RESULT 1  
YC17\_HAEIN STANDARD; PRT; 913 AA.  
AC P45114;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Probable tonB-dependent receptor H11217 precursor.  
GN H11217.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus.  
OX NCBI\_TaxID=727;  
[1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,  
Venter J.C.;  
RT "Whole-genome random sequencing and assembly of Haemophilus  
RT Influenzae Rd.,"  
RL Science 269:496-512(1995).  
[2]  
RP IDENTIFICATION BY MASS SPECTROMETRY.  
RX MEDLINE=20137488; PubMed=10675023;  
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,  
RA Gray C., Fountoulakis M.;  
RT "Two-dimensional map of the proteome of Haemophilus influenzae.,"  
RL Electrophoresis 21:411-429(2000).  
CC -|- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.  
CC -|- SUBCELLULAR LOCATION: Outer membrane (Potential).  
CC -|- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
-----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
CC or send an email to license@isb-sib.ch).  
-----  
CC EMBL; U32801; AAC22870.1; -  
DR TIGR; H11217; -  
DR InterPro; IPR000531; TonB\_boxC.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
DR Outer membrane; Receptor; Signal; TonB box; Complete proteome.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 913 PROBABLE TONB-DEPENDENT RECEPTOR H11217.

FT SITE 896 913 TONB C-TERMINAL BOX.  
SQ SEQUENCE 913 AA; 102768 MW; 01BEFDE5D6AEF617 CRC64;

Query Match 51.28; Score 2505; DB 1; Length 913;  
Best Local Similarity 51.58; Pred. No. 3.7e-149;  
Matches 495; Conservative 150; Mismatches 228; Indels 88; Gaps 16;

QY 1 MRSSFLRLKPCIFYL---MGVTLYHYSAEDAGRAGSEAQIQLVLDVHVAKRVKPKKKVF 57  
DB 1 MKKAIKLNLITLGINIGMTITQAQAEETLG-----QIDVVEKV-----ISNDKKPF 48

QY 58 TDARAVSTRQDIFKSENLENLIVRSIPGAFQTQDKSSGIVSLNIRGSGFRVNTWVDGI 117  
DB 49 TEAKAKSTRENVRKETQIDQVIRSIPIGAFQTQDKSGVSVNIRGENGLGRVNTWVDGV 108

QY 118 TQTFYSTDAGRAGSGSQFASVDSNFIAGLDVVKVSGSAGINLAGSANLRTLCVD 177  
DB 109 TQTFYSTALDSGQSGSQFGAAIDPFIAGVDVKNFSGAGINLAGSANFRTILGVN 168

QY 178 DVQGNNTYGLLLKGLTGTSTKGNMAAIGARKWLESGASGVLYGHRSRVAOYRVG 237  
DB 169 DVITDDPFGIILKMTGTSNATKSNFTMAAGRWLDNGGVYGVYGSQREVSQDRI- 227

QY 238 GGGQHIGNFGAEYLERRKQRYFVOEGALKFNSDGKWERDLQROQWK-YRKYKNYNN--- 293  
DB 228 GGBERLASLGQDILAKEAYFRNAGYIL--NPEGQWTPDLSKHKHSCNKPQYKNGDCS 285

QY 294 -----QELQK---VIEGHDKSWRENLAPOYDITIDP 322  
DB 286 YRIGTSAKTRREILQELLNGKKPKDIEKLQKNGDIEETDKSFERN-KDQYSVAPIEP 344

QY 323 SSLKQSQAGNLFKLEYDGVFNKYTAQFRDLNLTGKISKIIRNRYQFNGLSLNPTYNL 382  
DB 345 GSLQSRSHLLKFEYGDHQNLAQLRTLNDKIGSKRIENRYQVYNNNSYLDNL 404

QY 383 TAAVNSGRQYKPKSGKTGWLGLKDFFTYNNAKLTLDLNTATFRLPRELQTLTGFNYF 442  
DB 405 MAAHNIGKTIYKPGGFFAGVQADKLTKNVANIVDINNSHTFLPREIDLTGFNYF 464

QY 443 HNEYGKRRFPEELGFLFDGPDQDNGLYS--LGRFKDGKLLPKOKSTIVOPAGSOYNTF 500  
DB 465 TNEYSKRRFPEELSLFNDASHDGLSHSKRGYSTKLLPQSRVILQPSGKQKFTV 524

QY 501 YFDAALKKDIYRLNYSNTYTYRFGGTYGYGSDDEKFAFGENSPTYKKHKNRSGIY 560  
DB 525 YFDPAKSGIYHLNYSVNYETHAFNPEYGVY-----ENTAGQO-----IN 564

QY 561 EPVLKTKGKRANNSHVSISADFGDYFMPFASYSRTHRMPNIQDMYFSQIGDSCVHTALK 620  
DB 565 EPIHKSQGHKKAFAKNSATLSAELSDYFMPFTYSTRHMPNIQDMYFSQVSNAGVNTALK 624

QY 621 PERANTWQFGNTYKGLLQDDPFLGLKLVGRSIRIDNYIHNVYKWKMDLNGDIPSWVSS 680  
DB 625 PEQSDTTLQGLFNTYKGLFTQDDVLGVKLVGRSIFIKNYIHNVYGVWV--RDGMPYWAES 682

QY 681 TGLAYTQHNRKDKVKHGHGFELELNLDYGRFFTNLSYAYQKSTQPTNFSASPSNNAS 740  
DB 683 NGPKYTIHQNYKPIVKKSGVELEINYDMGREFANVSAYQRTNQPTNYADASPRPNAS 742

QY 741 KEDOLKQGYGLSRVSALPRDYGRLEVGTWRGLNKLTLGGAMRYFGKSIATAERYIDGT 800  
DB 743 QEDILKQGYGLSRVSMPLPKDYGRLELWRFDQKLTGLLAARYYKSKRATIEEYINGS 802

QY 801 NGGNTSNFROLGKRSIKQETLAPROLIFDYAAEYEPKKNLIFRAEVKNLFDRLYIDPLD 860  
DB 803 R-FKNTLRNRYTAVAKTDEIKQPIILDHLVSYEPDKOLIITIKAEVQNLDKRYVDPLD 861

QY 861 AGNDAATQRYSSFPDRDKDEDTYCNADKTLCLNGKYGGTSKSVLTPNFARGRTFLITMSYK 920  
DB 862 AGNDAASQRYSSIL-----NNSIECAQDSSAC---GSGDKTVLYNFPARGRTYLSLNYK 912

QY 921 F 921

Db 913 F 913

RESULT 2  
HXCL\_HAEIN  
ID HXCL\_HAEIN STANDARD; PRT; 744 AA.  
AC P44523;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 01-NOV-1995 (Rel. 32, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Heme/hemopexin utilization protein C precursor.  
GN HXUC OR HI0113.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus  
OX NCBI\_TaxID=727;  
RN [1]  
RC SEQUENCE FROM N.A.  
RC STRAIN=RD / KW20 / ATCC 51907;  
RX MEDLINE=95350630; PubMed=7542800;  
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,  
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,  
RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,  
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,  
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,  
RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,  
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,  
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.W., Smith H.O.,  
RA Venter J.C.;  
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae  
RT Rd.";  
RL Science 269:496-512(1995).  
CC -1- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW  
CONCENTRATIONS (BY SIMILARITY).  
CC -1- SUBCELLULAR LOCATION: Outer membrane (By similarity).  
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC -1- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE B OF  
H. INFLUENZAE.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
between the Swiss Institute of Bioinformatics and the EMBL Outstation -  
the European Bioinformatics Institute. There are no restrictions on its  
use by non-profit institutions as long as its content is in no way  
modified and this statement is not removed. Usage by and for commercial  
entities requires a license agreement (See <http://www.isb-sib.ch/announce/>  
or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; U32696; AAC21789.1; -.  
DR TIGR; HI0113; -.  
DR InterPro; IPR000531; TonB\_boxC.  
DR Pfam; PF00593; TonB\_boxC; 1.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KW Outer membrane; Transport; TonB box; Signal; Receptor;  
KW Complete proteome.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 744 HEME/HEMOPEXIN UTILIZATION PROTEIN C.  
SQ SEQUENCE 744 AA; 85043 MW; 74D94F72E41AEC31 CRC64;

Query Match 5.6%; Score 273.5; DB 1; Length 744;  
Best Local Similarity 20.9%; Pred. No. 9.7e-10;  
Matches 181; Conservative 136; Mismatches 366; Indels 183; Gaps 41;

QY 14 LMGTLYHYSAEDAGRAGSEAQIQLVLDVHVAKRVKPKKFTDARAVSTRQDIEK-S 72  
DB 15 LVGIQL-NVTAKONSSNRENLPIIVNTNDSNKL-----GRSLKQKNIEQXQ 65

QY 73 SENLDNIVRSIPGAFQTQDKSSGIVSLNIRGSGFRVNTWVDGITOTFYSTSTADGAG 132  
DB 66 ADNAANLINILPGVNWAGGFRPGQTLINMGDAEDRVQLDGAATKSFKEK----- 116

```
QY 133 GSSQFGASVDSNFTAGLDVYKSGSGSAGINSLAGSANLRLTLGVDDVVOGNNYTGILLKG 192
Db 117 -YQGSIFPELRLRRVTVDKNGSPQVGGFAGTVKFKTDARDFLQENQKIGGFLK- 174
QY 193 LTGNS-----TKGNMAAIGARKWLESAGSVGLYGHRSRVAQNYRVGGGGHIGNFG 247
Db 175 -YGNNSNNOKTYSTALVQLQNEKNIDL-----LLFGSVRN--AGDYKRPDMSKIL----- 222
QY 248 AEYLERRKQRYFVOEGALKFNDSGKWERDLQROOWKYKPKYNNNOELQYIGHDKSW 307
Db 223 ---FSKNQK---TGLIKLN-----WQISP-----EHLTLSSVYGIHKGW 257
QY 308 RENLAPQYDIPIPDSSLLKQOSAGNLKFLVDGVFNKYTAQFRDLNLTIGSKRKIINRYQ 367
Db 258 -EPAARKDIIP-----KPSLDIIRYGTD-TAWRKLVIYD-----QKDNITYLK 301
QY 368 FNYGLSLNPYNLNLTAAYNSGRQK--YPKGSKFTGMGLL--KDFEYVNNAKILDNNTA 423
Db 302 YNYPENNPWNLSTQFSYKTTQNDMRPKKASSGLVSLGNQSWITYSDL-TFDINNTS 360
QY 424 TFRLPRETELQTTLGFNYFHNKRPPELGLFFDGDQDNGLYSLVGRFKGDKGLLP 483
Db 361 TFNI-KTTVHELLFGLQWLNKT--RNT-----LMYDKSKVRKADYNY-GYFQ----- 403
QY 484 QKSTIVOPAGSQYFNTFYFDAALK-KDIYRLNYSTNTVGYRFGGEYTYGYSDDDEFKRAF 542
Db 404 ---PYMPSGRQYTAQFYLOQIWKNNII---FST---GVRY-----DHINNIG 443
QY 543 GENSPYKHKHNRCSGIYEPVLKYGKKRANNHS--VSISADFGDYFMPFASYSRTHRMP 600
Db 444 QKNLAL--KYNDISAG-----HDYSQKNYNGWSYVGLNVDVNHLSLFTNFSKTWRAP 495
QY 601 NIQEMYSQIGDSGVHTA---LKERANTWQGFNTYKGLLKQDDTLGLKLGVRSRID 657
Db 496 VIDEOYETQFOKSSVPATSLNLEKEMINQTRVGGIITLHNFQENDAPQFETTYFNRGK 555
QY 658 NYIH-----NYVGKWLNDLNGDI-PSWSSSTGLAVTIOHRNEKDKVHKHGFLELNLYDGR 711
Db 556 NEIFKTRGVNCVGNADNNKVCPIIEN-----YRNLPGYVIO-GAELEYQSTY 606
QY 712 FFTNLAYAQSTQTPNFSASESPNNASKEDQLKQGYLSRVSAALPRDYGRLEVGTRWL 771
Db 607 LFGEITYSVYKGRDT-----SPRN-----PWGKTSWIAEIPPKATTALGFNVP 652
QY 772 GNKLTGAMRYFGKSTRATAERYIDCTNGCNTSNFQLGKRSLKSTQETLARQPLPDF 831
Db 653 KYITLVGWRAEFVRQRDRSPL-----SGDPKASSWSLPASRGYS-----LHNL 695
QY 832 YAAYPEK--KNLIFRAEVKNLFDRLRY 855
Db 696 FLSWSPAKIKGMVNIITVDNLFENRAY 721

RESULT 3
TBPL_NEIGO
ID TBPL_NEIGO STANDARD; PRT; 915 AA.
AC Q01996;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPL.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID:485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA19.
RA MEDLINE=92394880; PubMed=1325963;
RA Cornelissen C.N., Biswas G.D., Tsai J., Paruchuri D.K.,
RA Thompson S.A., Sparling P.F.;
RT "Gonococcal transferrin-binding protein 1 is required for transferrin
utilization and is homologous to TonB-dependent outer membrane
```

```
receptors.";
RL J. Bacteriol. 174:5788-5797(1992).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
TRANSFERRIN UTILIZATION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
the European Bioinformatics Institute. There are no restrictions on its
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (See http://www.isb-sib.ch/announce/
or send an email to license@isb-sib.ch).
CC
CC EMBL; M96731; AAA25503.1;
DR PTR; A43335; A43335.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC. 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 24
FT CHAIN 25 915 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 38 45 TONB_BOX.
FT SITE 898 915 TONB_C-TERMINAL_BOX.
FT SITE
SQ SEQUENCE 915 AA; 102213 MW; 697CF74B1010422F CRC64;

Query Match 5.6%; Score 273.5; DB 1; Length 915;
Best Local Similarity 21.3%; Pred. No. 1.3e-09;
Matches 222; Conservative 129; Mismatches 327; Indels 365; Gaps 57;

QY 5 FRLPICPYLMGVTLHYHYAEDAGRAGEAQIQVLEVDHVHAK--RVPKDKKVTTDARA 62
Db 7 FRLNITLCLMTAL---PAYAENV-QAG-QAQEKOLDIQVAKKQKRRDNEVTGLGKL 61
QY 63 VSTRQDIPKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFRVNTWVDGIT 118
Db 62 VKTADTLKEQVLDIRDLTRDPGIAVVEQGRGASSG---YSIRG-MDKNRVSLVVDGLA 117
QY 119 Q-TFYSTSTDAG--RAGSSQFGASVDSNFTAGLDVYKSGSGSAGINSLAGSANLRLTIG 175
Db 118 QIQSYTAQAALGGTFTAGSSGAINIEIENVKAVEISKGSNSVEOGSAGALAGSAVFQKT 177
QY 176 VDDVVOGNNYTGILLKGLTGTNSTKGNMAAIGARKWLESAGSVGLYGHRSRVAQNYR 235
Db 178 ADDVIGEGRWGI-----QSKTAYSGKNRGL-----TQSIALAGR 212
QY 236 VGGG-----GOHIGNFGAEYLERKORYFVQEGALKFN-----SDSGKWERDLQROOW 283
Db 213 IGGAEALLIRTPGRHAGEIRAEHAGR-----GVQSFNRLAPVDDGSKYAFVEEB- 263
QY 284 KYKPYKYNNOEL-----QKYLEGHDK---SWRENLAQYDITPIDPSLKKQS-----AG 331
Db 264 ---CKNGGHEKCRANKPKDVVEDKROTVSTRDYTGPNRFLA--DPLSYESRSWLFRPG 317
QY 332 NLF--KLEYDGVFNKYTAQ---FRDLNTYKIGSRKII-----NNRYQFN--YG 371
Db 318 FRFENKRHYIGGILLERTQOTEDTRDMTPAFILTKRAVDANOKQAGSLRGNGKYAGNHRYG 377
QY 372 LSLNPYTNLNTAAYNSGRQYKPGSKFTGMGLLCKDFETYNNAKILDLNNTATFRLPRET 431
Db 378 -----GLFTSGENNAAPGAAY-GTGVFYD-----ET 402
QY 432 ELQTTLGFNYFHNKRPPEELGLFFD--GPDDQDNL----- 468
Db 403 HTKRYGLEIYVYTRADKDTWADYARLSYDRQIGLDNHFQOQTHCSADGSKYCRPSADKP 462
QY 469 YSYLGRFKGDKGLLPQKSTIVQAGSQYNTFYFDAALKKDIYRLNYSTNTVGY-RFGG- 526
Db 463 PSY---YKSDRVIVGESHKLLQAAFKKSFDTAKI-----RHNLNVN-LGYDRFGSN 509
```

QY 527 ----EYTGYYGSDDEKRAFGNSP----- 547  
 Db 510 LRHODY--YIGSAN--RAYSLKTPPNNGKKTSPNGREKNPYWYISGRGNVTVTRQICLP 564  
 QY 548 ---TYKKHCNRSC---GIYEPVLKK-----YKKRANNHVS-----I 579  
 Db 565 GNNTYTDCTPRISNGKSYAAVRDNVRLGRWADVGAGLRYDYRSTHSDGSGVSTGTHRTL 624  
 QY 580 SADFGDYFMPASYSRTH-----RMPNIQEMYSQIGDSGVHTALPERANTWQFG--- 630  
 Db 625 SWNAGIVLKPADWLDLRYSTGRFLPSFAEMYGWRSGDKIKAVKIDPEKSFNKEAGIVF 684  
 QY 631 -----FNTYKKGLLKQDDTLGLKLVGYRSRDN-----YIH----- 661  
 Db 685 KGDFGNLEASWNNAYRDLIVR-----GYEAQIKDGKEQVKGPNAYLNAQSARITG 735  
 QY 662 -NYGK--W---WDLNGDIP-SWVSSTGLAYT-IQHRNFQKDKVHKHGFEELELNVDYGRFF 713  
 Db 736 INILGKIDWGVND---KLPEGWYST--FAYNRVVRDIKRRADRTDIQSHL-FD----- 784  
 QY 714 TNLVYQKSTOPTNFSDASESPNASKEDQKOGYGLSRVSALPRDYGRLEVGTRWLN 773  
 Db 785 -----AIQPSRYV-----VGSYDQPEGKGVNGMLT-----YSRAKEITELLS 824  
 QY 774 KLTGGAMRYFGKSIKRAEERYIDGTNGGNTSNFROLGKRISQKOTETLARQPLIDFYA 833  
 Db 825 RALLNG-----NSRNTKATARRTRPWYIVDVSG 852  
 QY 834 AYEPPKNLIPRAEYKNLFDRIYI 856  
 Db 853 YTVVKHFTLRAGYVNLNHRIV 875

## RESULT 4

TB12\_NEIMB  
 ID TB12\_NEIMB STANDARD; PRT; 908 AA.  
 AC Q06987;  
 DT 01-FEB-1995 (Rel. 31, Created)  
 DT 01-FEB-1995 (Rel. 31, Last sequence update)  
 DT 16-OCT-2001 (Rel. 40, Last annotation update)  
 DE Transferrin-binding protein 1 precursor.  
 GN Tbp1  
 OS Neisseria meningitidis (serogroup B).  
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
 OX NCBI\_TaxID=491;  
 RN [1]  
 RC SEQUENCE FROM N.A., AND SEQUENCE OF 25-42.  
 RC STRAIN=CCUG 37603 / B16B6 / SEROGROUP B / SEROTYPE 2A;  
 RX MEDLINE=93345825; PubMed=8344530;  
 RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,  
 RA Jacobs E., Schryvers A.B.;  
 RT "Cloning and characterization of Neisseria meningitidis genes  
 encoding the transferrin-binding proteins Tbp1 and Tbp2.";  
 RL Gene 130:73-80(1993).  
 RP [2]  
 RN SEQUENCE OF 25-45.  
 RC STRAIN=CCUG 37603 / B16B6 / SEROGROUP B / SEROTYPE 2A;  
 RX MEDLINE=93307625; PubMed=8319886;  
 RA Griffiths E., Stevenson P., Byfield P., Ala'Aldeen D.A.,  
 RA Borriello S.P., Holland J., Parsons T., Williams P.;  
 RT "Antigenic relationships of transferrin-binding proteins from  
 Neisseria meningitidis, N. gonorrhoeae and Haemophilus influenzae:  
 RT cross-reactivity of antibodies to NH2-terminal peptides.";  
 RL FEMS Microbiol. Lett. 109:85-91(1993).  
 CC -1- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR  
 TRANSFERRIN UTILIZATION.  
 CC -1- SUBCELLULAR LOCATION: Outer membrane.  
 CC -1- INDUCTION: BY IRON STARVATION.  
 CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
 CC  
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
 CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way  
 CC modified and this statement is not removed. Usage by and for commercial  
 CC entities requires a license agreement (See http://www.isb-sib.ch/announce/  
 CC or send an email to license@isb-sib.ch).  
 CC  
 CC EMBL; Z15129; CAA78831.1; -  
 DR PIR; JN0819; JN0819.  
 DR PIR; S33154; S33154.  
 DR InterPro; IPR000531; TonB\_boxC.  
 DR Pfam; PF00593; TonB\_boxC; 1.  
 DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; 1.  
 DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
 KW Outer membrane; Receptor; Signal; TonB box.  
 FT SIGNAL 1 24  
 FT CHAIN 25 908 TRANSFERRIN-BINDING PROTEIN 1.  
 FT SITE 38 45 TONB BOX.  
 FT SITE 891 908 TONB C-TERMINAL BOX.  
 SQ SEQUENCE 908 AA; 101583 MW; FE2FF4974CAC3C31 CRC64;  
 Query Match 5.5%; Score 269; DB 1; Length 908;  
 Best Local Similarity 21.2%; Pred. No. 2.4e-09;  
 Matches 220; Conservative 119; Mismatches 335; Indels 366; Gaps 53;  
 QY 5 FRLKPCIFLYMLGVTLYHYSYAEDAGRAGSEAIQVLEDVHVAK--RVPKDKKVFDTDARA 62  
 Db 7 FRNLILCLSLMTALPV---YAENV--QAEQAQEKQLDTIQVAKKQKTRDRNEVTGLGKL 61  
 QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQDQK--SSGIVSLNIRGDSGFRVNTWVGIT 118  
 Db 62 VKSSDTLSKEQVLNIRDTRYDPGIAVVEQGRGASSG---YSIRG-MDKNRVSLTVDGVS 117  
 QY 119 Q-TFYSTSTDAG--RAGSSQFGASVDNFTAGLDVVYKGSFGSAGINSLAGSANLRTG 175  
 Db 118 QIOSYTAQAALGTRTAGSSGAINIEYENVKAVEISKGSNSSEYNGALAGSVAFOQT 177  
 QY 176 VDDVQGNNTYGLLLK-GLTGTNSTKGNMAAIGARKWLESAGSVGLYGHRSRVAQNY 234  
 Db 178 AADIIGEKQWGIQSTAYSGDKHALTQSLALAGR---SGGAELLITYTKRRGREHAH 233  
 QY 235 RVGGGQGHIGNFGAEXLERKQ-----RYFVQEGAL-----KFNSD-SGKWER-- 276  
 Db 234 KDAGKG--VOSFNRLVDEDKEGGSQYRYFIVEECHNGYAAACKNKLKEDASVKDERKT 291  
 QY 277 -----DLQOQWKYK-----YKNYNNQELQKYLEIGHDKSWRENLAPO 314  
 Db 292 VSTQDYTGNSRLLANPLEYGSQWLFPRGWHLNDRHYVGAVLERTQQTFTD--RDMTVPA 349  
 QY 315 YDITPID--PSSLK-----QQSAGNLFKLEY-DGVFNKYTAQPRDLNTK 355  
 Db 350 Y-FTSEYVPGSLKGLGKYSKDNKAERLFVQGEGSTLQIGIGYGVF-----YDERHTK 402  
 QY 356 IGRKIINRNYQFNYGLSLNPYTNLNTAAYNSGRQKYPKSGKFTGWGLLKDPEFYNNAK 415  
 Db 403 -----NR-----YGVVEY-VYHNADKDTWADYARLSYDRQG----- 431  
 QY 416 ILDNNTATFRLPRETELQTTLGFNYFNHNEYKGNRFPFELGLFDPDQNGLYSYLGRF 475  
 Db 432 -IDLN-----RLQOT---HCSHDSGDKNCRP-----DG---NKPYSF---Y 463  
 QY 476 KDGKGLLPQKSTIVQAGSOFYNT-----FYFDAAL----- 506  
 Db 464 KSDRMIEESRNLFQAVFKKAFDTAKIRHNLSINLGLDREKFSQLSHSDYLLQNAVAQYDL 523  
 QY 507 -----KKDIYRLVNSTVVG---YRFGGE-YT-----GYGSDDEEK 539  
 Db 524 ITPKPPFPNGSKDNPYRVSIGKTTVTNTPICRFGNNTYTDCTPRNIGGNGYAAVDNV 583  
 QY 540 RAFGENSPYKKHCNRSCGIEPYVLKYYKKRANNHVS-----ISADFGDYFMPFASY 593  
 Db 584 R-----LGRWADVGAGIRYDYRSTHSEKDSVSTGTHRNLSWNAAGVVLKPTWM 631  
 QY 594 SRTH-----RMPNIQEMYSQIGDSGVHTALPERANTWQFG-----FNT 633

```

Db 632 DLTVRSTGTFPSFAEMYGWRAGESLKTLDLKEKSFNREAGIVFKDGFNLEASFNN 691
QY 634 YKGLLKQDDTLGLKLVGYSRIDN-----YTH-----NVYK--WMDLNG 672
Db 692 AYRDLI-----AFGYETRTQNGOTSASGDGPGYRNAQNAIAGINILGKIDWHGVMG 742
QY 673 DIPSWVSTGLAY-----TIOHRNFKDKVHKHGFPELELYDY--GRFTNL 716
Db 743 GLPDGLYST-LAYNRKVKDADIRADFTVTSYLFDAVQPSRYVLGGLGYDHPDGIWGIN 801
QY 717 SYAYOKSTQPTNFSASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWRMLGNKLT 776
Db 802 MFTYSK-----AKSVDEL-----LGSQAL 820
QY 777 LGGAMRYFGKSIRATAERYIDGTNGNTSNFROLGRSRKQTTTLARQLIFDFYAYE 836
Db 821 LNG-----NANAKAASRR-----TRPVYVTVDSGYN 848
QY 837 PKKNLIFRAEVKNLDFRRYI 856
Db 849 IKKHLTLRAGYNLLNRYV 868

RESULT 5
TB1L_NEIMB STANDARD; PRT; 911 AA.
AC Q09056;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPL.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=CCUG 37608 / M982 / SEROGROUP B / SEROTYPE 9;
RX MEDLINE=93345825; PubMed=8344530;
RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,
  Jacobs E., Schryvers A.B.;
RT "Cloning and characterization of Neisseria meningitidis genes
  encoding the transferrin-binding proteins Tbp1 and Tbp2."
RL Gene 130:73-80(1993).
CC -1- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
  TRANSFERRIN UTILIZATION.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY IRON STARVATION.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
  between the Swiss Institute of Bioinformatics and the EMBL outstation -
  the European Bioinformatics Institute. There are no restrictions on its
  use by non-profit institutions as long as its content is in no way
  modified and this statement is not removed. Usage by and for commercial
  entities requires a license agreement (See http://www.isb-sib.ch/announce/
  or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z15130; CAA78833.1;
CC PIR; JN0821; JN0821.
CC PIR; S33156; S33156.
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_boxC; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
CC PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Receptor; Signal; TonB box.
KW SIGNAL
FT SIGNAL 1 24
FT CHAIN 25 911 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 38 45 TONB BOX.
FT SITE 894 911 TONB C-TERMINAL BOX.
FT SEQUENCE 911 AA; 101631 MW; 99283ABAE0B773E6 CRC64;

```

```

Query Match 5.2%; Score 256.5; DB 1; Length 911;
Best Local Similarity 21.6%; Pred. No. 1.5e-08;
Matches 220; Conservative 126; Mismatches 351; Indels 323; Gaps 55;

QY 5 FRLKPICFYLMGVTLHYHVSADAGRAGSEAOIQVLEDDVHVAK--RVPKDKKVFETDARA 62
Db 7 FRLNILCLSLMTAL--PAYAENV-QAG-QAOKQLDTIQVKAQKQKTRRNEVNTGLKL 61
QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQDK--SSGIVSLNIRGDSGFGRVMTVDGIT 118
Db 62 VKTADTLKSKEQVLDIRLDTRYDPCIIVVQCGRGASSG---YSIRG-MDKNRVSLTVDCGLA 117
QY 119 Q-TFYSYSTDAG--RAGSSQFGASVDSNFAGLDVVKVSGSGAGINSLAGSANLRPLG 175
Db 118 QIQSYTAQAALGGRTAGSSGAINIEIYENVKAVEISKGSNSVEQGGALAGSAVAFQTK 177
QY 176 VDDVVOGNNYTGLLGLTGTNSTKGNAMAAIGARKWLESGASVGVLVGHRRSRVAQNYR 235
Db 178 ADDVIGEGRWGIQSK-----TAYSGKNRGLTQSLA 208
QY 236 VGGGQHGHNFGAELERRKORYFVQEGALFNDSGKWERDLQR-----QOMKYKPY-- 288
Db 209 LAG---RIG--GAELLHTGR--RAGEIRAHEDAGRGVQSFNRKLPVEDSEVAYFIV 260
QY 289 -----KNYN--NOELOKYLECHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AGN 332
Db 261 EDECEGKNYETCKSKPKKVVVKDERQTVSTRDYTGPNRFLA--DPLSYESRSLFRPGF 318
QY 333 LF--KLEVDGVFNKYTAQ---FRDLNLTIGSKRKLIINRYQFNYGLSLNPNYLNLTAAYN 387
Db 319 RFENKRHYIGGLEHTQOTFTDRMTVPAPLTKAV-----FPAKSKQAGSL 364
QY 388 SGROKYPKSGKFTGWGLLKDFETYNNAKILDLNNTATPRLPRETELQTTLGFNFHNEYG 447
Db 365 PGCKYAGNHKY--GGLFTNGE--NGALVGAEGYGVFY--DETHTKSRYLEVYTNAD 418
QY 448 KNRPEELGLFFD--GPDQDNGL-----YSYLGKFKGDKGLLPQ 484
Db 419 KDTWADYARLSYDRQGIQGLDNHFOQTHCSADGSKYCRPSADKPFYS---YKSDRYVIGE 475
QY 485 KSTVQAPAGSYENTF-----YFDAALK-KDIYRLN-----YSTNT----- 519
Db 476 SHRLQAFAKSFDTAKIRHNLNVLGDFRDSNLRHODYYIQHANRAYSKTTPPKTANP 535
QY 520 -----VGYRFGGEYTYGSDDEFKRAFGENSPYTKKHCNRSC---GIYEPVLKK-- 566
Db 536 NGDKSKPYWVSIGGNNVVTGQI-----CLFGNN--TYTDCTPRSINGKSYAAVRDNV 586
QY 567 -----YKKKRANNHSVS-----ISADGDFYFMPFASYSRTH-----RM 599
Db 587 RLGRWADVAGLRDYDRSTHSDSGSVSTGTHRTLSNAGIVLKPADWLDLTYRTSTGPRL 646
QY 600 PNICEMYSQIGDSGVHTALKPERANTWQFG-----FNYYKGLLKQDDTL 645
Db 647 PSFEMTGNRSGVOSKAVKIDPEKSFNKEAGIVFKGDFGNLEASWFWNNAYKDLIVR---- 702
QY 646 GLKLVGYSRSDIN-----YIH-----NVYK--W--WDLNGDIP--SWYSS 680
Db 703 -----GYEAQIKNGKEAKGDPAYLNAQSAKITGINILGKIDWNGVMD--KLPEGWYST 754
QY 681 TGLAYTTOH--RNFKDKVHKHGFLELYDY---GRFTNLISYAYOKSTQPTNFSASESP 736
Db 755 --FAYNRVHVRIKIKRADRTDIQSHL--FDAIQPSRYVYVGLY----- 793
QY 737 NNASKEDQLKQGYGLSRVSALPRDYGRLEVGTWRMLGNKLTGGAMRYFGKSIRATAERY 796
Db 794 -----DQPEGKVGWNGMLT-----YSKAKETELGGRALLNG----- 826
QY 797 IDGTNGGNTSNFROLGRSRKQTTTLARQLIFDFYAYAEKPKKNLIFRAEVKNLDFRRYI 856
Db 827 -----NSRNTKATARRTPWYIVDVSGYITIKKHFTLRAGYNLLNRYV 871

```





Qy 562 P-----VLKYGKRRANNHSHSVISADFGDYMPFASYSRTHRM 599  
Db 675 PLPKNSNKPDEVKKNVQNDIYAKONKKYKAHSYFVSTIDTSLRLQLKYSKGFRT 734  
Qy 600 PNIQEMFYS-QIGDSGV--HTALKPERANTWQFGFNYYKKGLLKQDDTLGLKLVGYRSRI 656  
Db 735 PTSDMEYFTFKHPDFTLLPNAULKPELAKTKTAFT-----LHNDWGWGFISSLFKTNY 788  
Qy 657 DNYIHNYYGKWDLNGDIPSWSTGLAYTIOHRNFKVKHKGFELELNVDYGR-----711  
Db 789 KNFIDLIFKG-----EKDFPLVSGSSLPFLSYQINRDNASLKGIEINSKVFLGKMAKFM 844  
Qy 712 --FTFTLSYAYQKSTOPTNFSASESPNNASKEDQLKQGYGLSRVSAALPRDYGRLEVGT 769  
Db 845 DGFNLSKYTYQKGRMGNI-----PMNAIQPRTWYVGLGYDHPN-----884  
Qy 770 WLGNKLTGGMARVFGKSIRATAEERVIDGTNGTNSFRQLGKRSTIKQRET-----LARQ 825  
Db 885 -----HKFGDFVTY-----HVASKNPEDTYDIYAKDK-----KOTDTSIKWRKS 925  
Qy 826 PLIFDYAAVEPKNLIKRAEVKNLFRDRYIDPLDAGNDAATORYYSSFPKPKDDEVTVC 885  
Db 926 YTIIDLLGYVQPIKNTIRAGVYNTNRKYIT-----WDSA--RSIRSGTNSVIEQ---975  
Qy 886 NADKTLNCGYGTGSKSVLNFARGRTFLTMSYKF 921  
Db 976 -----STGLGINRFYAPGRNRYKMSVQFEE 999  
RESULT 8  
ID LBPA\_NEIMB STANDARD; PRT; 943 AA.  
AC Q06379; Q9JYK5;  
DT 01-NOV-1995 (Rel. 32, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DT 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Lactoferrin binding protein A precursor (Iron-regulated outer membrane protein A).  
GN LBPA OR IROA OR NMB1540.  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=BNVC / SEROGROUP B;  
RX MEDLINE=94011384; PubMed=8406871;  
RA Pettersson A., van der Ley P., Poolman J.T., Tommassen J.;  
RT "Molecular characterization of the 98-kilodalton iron-regulated outer membrane protein of Neisseria meningitidis";  
RL Infect. Immun. 61:4724-4733(1993).  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC58 / SEROGROUP B;  
RX MEDLINE=20175755; PubMed=10710307;  
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Clecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B., Cotton M.D., Uitterback T.R., Khouri H., Qin H., Vamathevan J., Gill J.J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58";  
RL Science 287:1809-1815(2000).  
CC -!- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- INDUCTION: BY IRON STARVATION.  
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC EMBL; X69214; CAA49148.1; -  
DR EMBL; AE002504; AAF41895.1; -  
DR TIGR; NMB1540; -  
DR InterPro: IPR000531; TonB\_boxC.  
DR Pfam: PF00593; TonB\_boxC.1  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;  
KW Complete proteome.  
FT SIGNAL 1 27 POTENTIAL.  
FT CHAIN 28 943 LACTOFERRIN BINDING PROTEIN A.  
FT SITE 826 943 TONB C-TERMINAL BOX.  
FT CONFLICT 8 8 P -> Q (IN REF. 1).  
FT CONFLICT 16 18 IAT -> VAA (IN REF. 1).  
FT CONFLICT 22 22 A -> S (IN REF. 1).  
FT CONFLICT 26 31 OAGGAT -> NPETA (IN REF. 1).  
FT CONFLICT 43 43 I -> V (IN REF. 1).  
FT CONFLICT 64 64 V -> A (IN REF. 1).  
FT CONFLICT 233 233 R -> H (IN REF. 1).  
FT CONFLICT 243 243 E -> A (IN REF. 1).  
FT CONFLICT 247 247 D -> N (IN REF. 1).  
FT CONFLICT 257 269 DIKRTREPFVS -> GIKKPSGEGYFLA (IN REF. 1).  
FT CONFLICT 273 275 RES -> SEL (IN REF. 1).  
FT CONFLICT 281 281 L -> V (IN REF. 1).  
FT CONFLICT 286 286 YGK -> NGN (IN REF. 1).  
FT CONFLICT 313 313 Q -> M (IN REF. 1).  
FT CONFLICT 389 389 E -> K (IN REF. 1).  
FT CONFLICT 420 423 KNLV -> OKLI (IN REF. 1).  
FT CONFLICT 455 455 A -> K (IN REF. 1).  
FT CONFLICT 546 546 K -> N (IN REF. 1).  
FT CONFLICT 564 572 STGFENNQ -> YSDYTKG (IN REF. 1).  
FT CONFLICT 658 658 L -> V (IN REF. 1).  
FT CONFLICT 667 667 V -> L (IN REF. 1).  
SQ SEQUENCE 943 AA; 105680 MW; BD569ECACFC01A8A CRC64;  
Query Match 5.0%; Score 242.5; DB 1; Length 943;  
Best Local Similarity 20.2%; Pred. No. 1.1e-07;  
Matches 222; Conservative 149; Mismatches 380; Indels 347; Gaps 56;  
Qy 9 PICFYLGMVTLHYHVAEDAGRAGSE-AQIQVLEVDVHVAKRVPKDKKVTVDARAVSTQ 67  
Db 8 PLTLTALAIATAPYAAQAAGATPDAAQTQSLKEITVRAAKVGRSKEATGL-----G 61  
Qy 68 DIFKSSNLD-----NIVRSIPGAFTQDDKSSGIVSLNIRGDSGFGVRVMTVDGITQ 119  
Db 62 KIVKTSNLTKEQVIGRLDTRDPGVAVVEQNGASGGYSIRGVDK-NRVAVSVQDVAQ 120  
Qy 120 ----TFYSTTDAGRAGSQFGASVDSNFIAGLDVYVVGSGFSAGINSIAGSANLTGL 175  
Db 121 IQAFTVQGLSGYGGGSGGAIN-EIEYENISVTEIDKAGSGSDHSGALGGAFAFTKE 179  
Qy 176 VDDVQGNNTYGLLLKGLTGTNSTKGNMAAIGA-----RKWLES-----G 216  
Db 180 AADLISDGKSGIOAK--TAYSKNRQFMKSLGAGFSKDGWEGLLIRTERQRETRPHGD 237  
Qy 217 ASGVLYGHSRRSAQNYRVYGGGQHGIFNGFAEYLFERRKOR--YFVOEGALKFNSSDGK 274  
Db 238 IADGVEYIDR-----LDAPQTYDIKRRTREFFSVEG-----271  
Qy 275 ERDLQEQMKYKPKYKNYNNOELQYIEGHDKSWRENLAPOYDITPIDPSLKKQOSAGNLF 334  
Db 272 ERE-SKPVAKLAGYGYKLNQLNRWYKER-----IEQNOPLSAEEE-----311  
Qy 335 KLEYDGVFNKYTAQFRLDNTKIGSRKLIINRYQFNGLSLNPTN---LNLTAAYNSGR 390  
Db 312 -----AQVREAQAR-----HENLSAQAYTGGGRILPDPMDYRSGS 346









```
QY 250 YLERKORYFVOEGALKFNDSOGKWERD-----LQOQKQYKPKYKNYNQEL 296
D 248 RLDAFRQTYDIQONKKAAYELABGEPKPVAKLAGNGYLKQLNKRWVEERKKNQPL 307
QY 297 QKYIEG---HDKSWRENLAPOYDIT---PIDPSSLKOOSAGNLFKLEY-----DGVF 342
D 308 NAEEMVREAAQHENSQAQ-SYTGGRILPDPMYRSGSLWAKLGYRFGGRHYVGVE 366
QY 343 --NKYTAQFRDLNFK---IGSRKIINRN---YQFNGLSNPY-----TNLNL 383
D 367 EDTQRYDIRDTEKQYGYGDEAKFKNSGOYDGNDFRGLFVFNIEBWKDGTNLVK 426
QY 384 AAYNSGRQKY-PKSGKTGWLKDFE-----TYNNKILDLNNT----- 422
D 427 IGLKYSRTKETDEHRRRRMGLLYRYENEAYSDNWADKAVLSFDKQGVATDNTKLNCA 486
QY 423 -----ATFRLP---RETELQTLGFNFHN 444
D 487 VYPAVDSKRASADKPKSYDSSDRFHYRQHNVLNASFEKSLKNKWKHHLTGLGFGYDAS 546
QY 445 E-----YGNRFPPEELGFLGDPDQDGLY-----SYLGRFKGDKGLLPQ 484
D 547 KAISRPQLSHNARISESTG--FDDNNQDKYLLGKPEVVEGSGVCIETLSRKP-CVPR 603
QY 485 K---STIVQAGSOYFNTFFDAALKKDIYRLNISTWTVGYRFGGEYTYGYGSDDEPKRA 541
D 604 KINGSNHISLINDRFSIGKYGFDFSLGGRYDRQNETTSEELVRSG----- 647
QY 542 FGENSEPTYKHCNRSCG---IYEPVLKKGKRRANNH-SVSIADFGDFMPFASYSRTH 597
D 648 -----RYVDRSWNSGIVFKP-----NRHFSVSYRASS-----F 676
QY 598 RMPNIOEY-----FSQIGDSGVHTALKPERANT-----WQ---FGF-----NTYKKGL 638
D 677 RTPSQBELGDIYHDYKQWQRPALKSEKAAANREIGLQWKGDFGLEISSFRNRYTDMI 736
QY 639 LKQDDTLGL-KLVGYRSRID--NYIH-----NYYGK--WDLNGDIPSWSSVSTGLA 684
D 737 AVADHKTLQLPDSTGRLETFEIDIRYNAQNMSLOGVNILGIDWNGVYKLPGLYTT-LA 795
QY 685 YT-----IQHNEFK-DKVHKGHFELELNYDY--GRFTNLSYAYKSTQPTNF 729
D 796 YNRIKPKSVSNRPDLRSYALDAVQPSRYVLGFGYDQPEGKWCANIMLYSKGNPDE- 854
QY 730 SDASESPNNASKEDQLKQGGYLSRVSALPRDYGRLEVGTRW-----LGNKLTLLG 779
D 855 -----LAYLAGDQKRYSTKRASSSWSTADVSAYLNLKRLTLRA 893
QY 780 AMRYFG-----KSIRATAERYIDGTNGNTSNFQPKRSIKQETTLARQPLIFDFY 832
D 894 AIYNIGNRYRVWTWESLQRTAES---TANRHGSDSNYGR-----Y 929
QY 833 AAYEPKKNLIFRAEVK 848
D 930 AA--PGRNFSIALEMK 943

RESULT 12
ID TBPI_HAEIN STANDARD; PRT; 912 AA.
AC P44970;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable transferrin-binding protein 1 precursor.
GN TBPA OR TBPI OR H10994.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
```

```
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.P.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RL Rd.";
RL Science 269:496-512(1995).
CC -|- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION (BY SIMILARITY).
CC -|- SUBCELLULAR LOCATION: Outer membrane (Potential).
CC -|- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32780; AAC22656.1; -
CC TIGR: H10994; -
CC InterPro: IPR000531; TonB_boxC.
CC Pfam: PF00593; TonB_boxC; 1.
CC PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
CC PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Receptor; Signal; TonB box; Complete proteome.
CC SIGNAL 1 23 POTENTIAL.
CC CHAIN 24 912 PROBABLE TRANSFERRIN-BINDING PROTEIN 1.
CC SITE 50 57 TONB BOX.
CC SITE 895 912 TONB C-TERMINAL BOX.
CC SEQUENCE 912 AA; 103233 MW; 294C08991A652CC2 CRC64;

Query Match 4.9%; Score 238.5; DB 1; Length 912;
Best Local Similarity 20.3%; Pred. No 2e-07;
Matches 215; Conservative 134; Mismatches 413; Indels 299; Gaps 50;

QY 5 FRUKPICFYLMG--VTLYHYSYAEDAGRAGEAQIQVLED-----VHVKAARV--PKDK 55
D 7 FRLSIISCLLISCVRAETQSIKDKTKEAISSEVDSTQSTDELETISVTAEKVRDRKDNE 66
QY 56 VFTDARAVSTRQDIFKSSSE-NLDNIVRSIPGAFTQODKSSGIYSLNIRGDSGFRVNTMV 114
D 67 VTGLGIKIITSEISIREQVLNIRDLTRYDPGISVVEQGRGASSGYSIRG-MDRNRVALLY 125
QY 115 DGITQT-FYSTSTDAGRAGSSQFGA--SVDSNFIAGLDVYVKGFSGAGINSLAGSANL 171
D 126 DGLPQTQSYVQSPVARSYSGTGAINEIYENYKRAVEISKGSSEYNGALAGSVTF 185
QY 172 RTLGVDVVOGNNYTLGLLKGLTGTNSTKG--NAMAAGARKWLESGASVGLYHSHRRS 229
D 186 QSKSAADILLEGDKSWGIQTKN-AYSSKNKGFTSLAVAG-----KQGGFGLAIYQTRNS 239
QY 230 VAQNYRVGG--GGQHIGNFGAEYLERKQRYFVQE----- 262
D 240 IETQVHKDALKGVSQSYNRLIAK--PENQSAIFYMEDECPKGYDECIPSAKPPAILSTKKE 297
QY 263 -----GALKFNSDSGKWERDLQROQWKYKPKYKNYNQELQKYTEGHDKSWRENLAPO 314
D 298 TVSVSYDTGANRIKPNPKMYE---SQSWFLRGYHFSEQH---YIGGFTEFTQQ----K 346
QY 315 YDITPID-PSSLKQSQSAGNLF-----KLEY-----DGVFNKYTA--OFRDLNFK--- 355
D 347 FDIRDNTFPAYLSPTKEKGLDANRPFPYKQDYGAYQHTEDGRGVKYASGLYFDEHRRKQRY 406
QY 356 -----IGSRKIINRNRYQFNGLSNPY---TNLNLTAAYNSGRQYKPGSKF 399
```

```
Db 407 GIEYIENKKNAGIIDKAVLSANQQ-----NIILDSYMRHTCSL-----YPNPSKN 453
QY 400 TGMGLLKDFETYNNAKILDLNNTATFRLPRETELQT-----TLGFNYF-----HNE 445
Db 454 CRPTLDKPYSYRSDRVNRYEKKHMLQLNLEKKIQIOWNLTHQIVENLGFDFDTSALQHKD 513
QY 446 YGNRPFEEGLGDFD--GPDQDGL--YSYLGKFGDKGLLPKOKSTIVQAGSQYFNFFY 501
Db 514 YLRRVIATADSIDKTGKTRRGLREYPYL-----YKPKPYF----- 552
QY 502 FDAALKKDIYRLNVTNTVGYRGEGVTGYGSDDEFKRA-----FGENSPY 549
Db 553 -----AGEDHCNTQGSNNYRDCVKRLIKGKNYFYFARNMAL 590
QY 550 KKHONRSCGIYEPVLKYGKRRANHSVSI-----SADEGDFYMPF-----ASRTH 597
Db 591 GKYYDLGLGIRYDV-----SRTKANESTISYGVKFKNFWSNNTGIVIKPTEWLDLSYRLSTGF 646
QY 598 RMPNIQMYFSQIG-----DSGVHTALKPERANTWQFGENTYKGLLKODDPLGLKLGYRS 654
Db 647 RNPFSFAEYMGWRYGKGNDEYVYGVKFKPETSNRQOEF-----GLALKGDFGNEISHFSN 699
QY 655 RIDNYIHNVYKGMWDLNGDIPSWSSYGLAYTIOHRNFKDKVHKHKGFELELNDYDYGREET 714
Db 700 AYRNLI--AFAEELSKNG-----TGKNGYGYHNAQNAKL--VGVNITQAQLDNFGLWK 747
QY 715 NLSYAYOKSTQPTNFSASESPNNASKEDQKQYGLSLRSYAL-----PRDY-----GRL 764
Db 748 RIPYGMV-----ATPAYNRVKYKQO--KINAGLASVSSYLFDAIQPSRYIIGLYD 796
QY 765 EVGTRWLGKLTGGAMRYFGKSRATAEERYIDGTNGNGNTSNFRQLGKRSIKOTETLAR 824
Db 797 HISNTW-----GVNATRTQS--KAKSQNELLGKRALGNS-----RDVKSTRKLTR 840
QY 825 QPLIEDFYAAEYKPKNLIFRAEYKKNLFDREYIDPLDAGNDAATQRYYSFDPKDKDEDVT 884
Db 841 AWHILDVSGYMKANKMLRLGLYNLFNRY-----VT 873
QY 885 CNADKTLGNGYKGTGSKSVLTNFAR-----GRTEFLITMSYKF 921
Db 874 WEAVRQTAQAV--NQHQNVGNTRYAASGRNYYLTILEMKF 912
RESULT 13
Y262_HAEIN STANDARD; PRT; 723 AA.
AC P44600;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable tonB-dependent receptor HI0262 precursor.
GN HI0262.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kralavag A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Sequence 269:496-512(1995).
```

```
Db 419 LFGEKLVSPSVRYDHYDTS-----SKTVYKDNHLSPATKLTWIVTNWLDL---TA 467
QY 592 SYSTRHPNIOEYFQIGDSGVHTA-----LKPERRANTWQFGENTY 634
Db 468 KYNEAFRAPSQMERFV-----SGAHFGANTGLDHNFRVANPNLRPETAKNKEITANLH 522
QY 635 KGLLKQDDTLGLKLVGRSRIIDNIH-NVY-----GKWDNLNGDIPSWVSTGLA 684
Db 523 FDSLFGKQDFKFIENYFRNDVKDFINLKIFNDAKTSASAGAPNTINGAL-----LP 574
QY 685 YTHQRNFKDKVHKHGFELNIDYGR--PFTNLSYAYQKSTOPTNFSDESPPNASK 742
Db 575 KNSOYQNTITN-ARLSGIELQAQYQTERLTFTNY-----GSTKG 612
QY 743 DQLKQGVGLSRVSALPRDYGRLEVGTRW--LGNKLTILGGAMRYFGKSIRATAERYIDGT 800
Db 613 KDKDSGEALSNIAA-----SKIGVGNVALVKDFVGATVTHAAQ-----654
QY 801 NGGNTSNFRLGKRSIKQETELARQPLIFDYAAEPK-----KNLIFRAEVKNLFDORRYI 856
Db 655 -----RRVPRKDHSTVTPSYILTLDRATYAPLKGEMKNLRLDPALENLFDKRYQ 702
QY 857 DPLD-----AGNDAATQRYYS 872
Db 703 PAFSLMEGTGRNAKISAVYS 722

RESULT 14
HGPA_HAEIN
ID HGPA_HAEIN STANDARD; PRT; 1077 AA.
AC G9ZA21; O9R649;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hemoglobin and hemoglobin-haptoglobin binding protein A precursor
DE (heme-repressible hemoglobin-binding protein) (Hgb).
GN HGPA.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=HI689 / Serotype B;
RX MEDLINE=99233576; PubMed=10220170;
RA Jin H., Ren Z., Whitby P.W., Morton D.J., Stull T.L.;
RT "Characterization of hgpa, a gene encoding a haemoglobin/haemoglobin-
RT haptoglobin-binding protein of Haemophilus influenzae.";
RL Microbiology 145:905-914(1999).
RN [2]
RN SEQUENCE OF 1-145 AND 988-1077 FROM N.A., AND PARTIAL SEQUENCE.
RP STRAIN=HI689 / Serotype B;
RX MEDLINE=96333350; PubMed=8757844;
RA Jin H., Ren Z., Pozsgay J.M., Elkins C., Whitby P.W., Morton D.J.,
RA Stull T.L.;
RT "Cloning of a DNA fragment encoding a heme-repressible hemoglobin-
RT binding outer membrane protein from Haemophilus influenzae.";
RL Infect. Immun. 64:3134-3141(1996).
RN [3]
RP ROLE OF CCAA NUCLEOTIDE REPEATS.
RX MEDLINE=99412292; PubMed=10482534;
RA Ren Z., Jin H., Whitby P.W., Morton D.J., Stull T.L.;
RT "Role of CCAA nucleotide repeats in regulation of hemoglobin and
RT hemoglobin-haptoglobin binding protein genes of Haemophilus
RT influenzae.";
RL J. Bacteriol. 181:5865-5870(1999).
RC CC -I- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE.
CC -I- SUBCELLULAR LOCATION: Outer membrane.
CC -I- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
```

```
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAINING.
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.
CC -I- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U51922; AAD10835.1;
CC InterPro: IPR000531; TonB_boxC.
CC Pfam: PF00593; TonB_boxC; 1.
CC PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Transport; TonB box; Multigene family; Signal;
CC Receptor; Repeat.
CC SIGNAL 1 24
CC CHAIN 25 1077 HEMOGLOBIN AND HEMOGLOBIN-HAPTOGLOBIN
CC
CC DOMAIN 26 69 11 X 4 AA TANDEM REPEATS OF P-T-N-Q.
CC REPEAT 26 29 1.
CC REPEAT 30 33 2.
CC REPEAT 34 37 3.
CC REPEAT 38 41 4.
CC REPEAT 42 45 5.
CC REPEAT 46 49 6.
CC REPEAT 50 53 7.
CC REPEAT 54 57 8.
CC REPEAT 58 61 9.
CC REPEAT 62 65 10.
CC REPEAT 66 69 11.
CC SITE 78 85 TONB BOX.
CC SITE 1060 1077 TONB C-TERMINAL BOX.
CC SEQUENCE 1077 AA; 122813 MW; 693F673BB5AC59F1 CRC64;

Query Match 4.7%; Score 228.5; DB 1; Length 1077;
Best Local Similarity 19.5%; Pred. No. 1e-06;
Matches 211; Conservative 131; Mismatches 351; Indels 389; Gaps 48;

QY 40 LEDVHVYKARVPKDKKVFTD-ARAVSTRQIDFK-SSENLDNIVRSIPGAFQODKSSGIV 97
Db 77 LEQINVSGSTENTDTRAPKIAETVKTAKLEQAQDVKDLVRYETGI-----TVV 138
QY 98 SLNIRGDSGF-----GRVNTWVDGITOTFYSTSTDAG-----RAGSSQFGASVDSN 144
Db 129 EAGRFNSGFVRGVVEENRVAVOIDLHQH--ETISQGFKELFEGYGNFNTRNRAEIE 186
QY 145 FTAGLDVVRKGSFSGSAGINSAGLSANLRTLGVDVVDVQGNNTYGLLLKLTGTNTSGNAM 204
Db 187 TLKQVTRIRKADSLKSGSGALGGSVSLDTKARDYLLNKNYASYKRGYNATDNQNLNTL 246
QY 205 AIGARKWLESAGSVGLYCHSRSSVAQNVYRGGGQHIGNFG-----AYELERRK 255
Db 247 TLGGRYKYFD--AIAVL--TSRK-----GHELENFGYKNYNDKIQKTRKAD 290
QY 256 QRYFVQEGAL-----KNSDSGKWERDLQRQWKY--KP--YKNYNNQELOKY 299
Db 291 PYRTODSALLKIGFQPTENHRFSVVADLYKQTSKGDHFSYTLKPNQTYWYDEKLR-- 348
QY 300 IECHDKSWRENLA-----POYDITPIDPSSLKQOS-----AGNLF 334
Db 349 -HTNDKVERKNIAFVYENFETFPFWDTLKITYSHQITTSARTDDYCDGNDKCALAGNPL 407
QY 335 KLEY-----DGVFNKYTAQFDRLN-TKIGSRKLIINRNQYFNGLSNPTNLT 384
Db 335 KLEY-----DGVFNKYTAQFDRLN-TKIGSRKLIINRNQYFNGLSNPTNLT 384
```

```
Db 408 GMYNODNQLVGKDG---KSAYQDINKTQVIKERL-----PFTKPN--- 446
QY 385 AYNSGRQYKPKSGKFTGWLK-----DFETYN----- 412
Db 447 ---GRWRFHK---VDWDLAKKKYGPVPIYASCLEEDNDPSEFCYEVKTKKENTFEI 498
QY 413 NAKILDNLNTA-----TFLPRET-----ELQTLIGFNYFHN----- 444
Db 499 NGARYDLLSEADKNVLSDEQRPLTNVSYLFSQDGLNCKTKTLGFKKRRNLLKIFLFEVI 558
QY 445 -----EYGNRPPEELGLEFDGPDQDNGLYSYLGRFKGDKGLL----- 482
Db 559 EKRCQYGRKTKVKAN-----DQLSGYLFMPNKKGYQANLWSQDLTSETKQINLDEL 610
QY 483 -----POKSTIVQAG-----SQY----- 496
Db 611 TKHLELGKTQHDLSYGLWSEMKSTNLAGDPTPLNWKWAAQYPHNCATFLPDPSTWTPNA 670
QY 497 -----FNTYFEDAAL-----KDIYRLNYSNTVGYRFG-----GE 527
Db 671 KPTLNPERTSTLCNNVNVFSFLIPVKTKGALYFINDRVNNYVAFNLGYRYDRVKYEPE 730
QY 528 Y-----TGYYSDDDEFKRAFGENSP--TYKKHCNRSCTGYPVLKKYKCKRA 572
Db 731 YIPGKTPKIPDDMTNLYIKTPEFDASKADSDPELSKKEANAANKETAQPK--KFS 788
QY 573 NNHSVISADFGYFMPFASYSSTHRMNPQIEMWFS-QIGDSGV--HTALKPRANTWOF 629
Db 789 SSYSFTGTLTDLNWLRLQAKYSGFRAPTSDEIYFTFKHPDFSIOPNRDLQPETAKTEL 848
QY 630 GFNTYKGLLKQDDTLGLKLVGYSRIDNKHNVYCKWDLNGD---IPSWYSSTGLAYT 686
Db 849 SLTVH-----NDMGYITTSFDTRYQNFIDLSYQGRDVGHSKLIPF----- 891
QY 687 IOHRNFKDKVHK-----HGFELELNLDYGR-----FTNLISYAYQKSTQPTNFSDAE 734
Db 892 ----HYQNVNRPNKAVTGFETASQISLGNITLFGNFSLSYKTYTKQGRINGNIPNAL 947
QY 735 SPNASKEDQKQYGLSRVSLPRDYGRLEVGRWLNKLTIGGAMRVFGKSIRATAEE 794
Db 948 QPRTAV-----YGVSYVH---PDDKYGLDL-----YISHASAKNAED 981
QY 795 RYIDGTNGGNTSNFROLGKRSIKQETLARQPLIEDFYAAYEPKKNLIPRAEVKNLFDRR 854
Db 982 TY-----NMFKYKGGTD-STIKWRSKSYTTIDLLGYKPIKNTLRAGVYNLTNRK 1032
QY 855 YI 856
Db 1033 YI 1034

RESULT 15
HGP2_HAEIN
ID HGP2_HAEIN STANDARD; PRT: 999 AA.
AC P44809;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable hemoglobin and hemoglobin-haptoglobin binding protein 2
DE precursor.
GN HI0661.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RC SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
```

```
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
RN [2]
RP REVISIONS.
RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.,
RA Peterson J., Hickey E., Dodson R., Gwinn M.;
RL Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=317488; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
RN [4]
RP CONCEPTUAL TRANSLATION.
RA Coudert E.;
RL Unpublished observations (AUG-2001).
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A WAY TO
CC AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; TWO FRAMESHIFTS WERE
CC INTRODUCED IN THE REPEATS REGION AND IN THE LEADER PEPTIDE TO
CC MAXIMIZE THE SIMILARITY WITH OTHER ORTHOLOGS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32749; AAC22319.1; ALT_SEQ.
CC TIGR: HI0661;
CC InterPro: IPR000531; TonB_boxC.
CC Pfam: PF00593; TonB_boxC; 1.
CC PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Transport; TonB box; Multigene family; Signal;
CC Receptor; Repeat; Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 999 PROBABLE HEMOGLOBIN AND HEMOGLOBIN-
FT HAPTOGLOBIN BINDING PROTEIN 2.
FT DOMAIN 26 53 7 X 4 AA TANDEM REPEATS OF Q-P-T-N.
FT REPEAT 26 29 1.
FT REPEAT 30 33 2.
FT REPEAT 34 37 3.
FT REPEAT 38 41 4.
FT REPEAT 42 45 5.
FT REPEAT 46 49 6.
FT REPEAT 50 53 7.
FT SITE 63 70 TONB BOX.
FT SITE 982 999 TONB C-TERMINAL BOX.
SQ SEQUENCE 999 AA; 114690 MW; 1A17AAB220092B7D CRC64;
```

```

Db      968  GTSNVIDQ-----: : |:
                                     |
                                     : : | | : : : : : |
                                     : : | | : : : : : |
-----STGLGINRFYAPGGRNYKMSVQFEF 999

Search completed: July 24, 2002, 09:00:22
Job time: 413 sec

```







GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:59:42 ; Search time 68.49 Seconds  
(without alignments)  
2326.301 Million cell updates/sec

Title: US-09-762-926-6  
Perfect score: 4894  
Sequence: 1 MRSSFLRPICFLMGVTLV.....SVLTNFARGRTFLTMSYKF 921

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL\_19:\*  
1: sp\_archaea:\*  
2: sp\_bacteria:\*  
3: sp\_fungi:\*  
4: sp\_human:\*  
5: sp\_invertebrate:\*  
6: sp\_mammal:\*  
7: sp\_mhc:\*  
8: sp\_organelle:\*  
9: sp\_phase:\*  
10: sp\_plant:\*  
11: sp\_rodent:\*  
12: sp\_virus:\*  
13: sp\_vertebrate:\*  
14: sp\_unclassified:\*  
15: sp\_rvirus:\*  
16: sp\_bacteriaph:\*  
17: sp\_archaeap:\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

#### SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	4883	99.8	921	16	Q9JR50 neisseria m
2	4763.5	97.3	922	16	Q9JTN4 neisseria m
3	1096.5	22.4	755	16	Q9PIV1 campylobact
4	901	18.4	753	2	Q93052 shigella fl
5	900	18.4	753	2	Q93354 escherichia
6	468.5	9.6	885	16	Q92NX4 rhizobium m
7	435.5	8.9	883	2	Q9XSP4 pseudomonas
8	435.5	8.9	891	16	Q9HYJ7 pseudomonas
9	428.5	8.8	989	16	Q9HUR6 pseudomonas
10	426	8.7	916	2	Q9RHT4 pseudomonas
11	412	8.4	899	2	P72473 serratia ma
12	357.5	7.3	848	16	Q9CKJ4 pasteurrella
13	346	7.1	747	16	Q98L70 rhizobium l
14	311	6.4	782	2	Q93SH4 bradyrhizob
15	309.5	6.3	784	16	Q9KLM6 vibrio chol
16	279	5.7	739	2	O87381 haemophilus

17	272.5	5.6	907	2	Q9KI54	Q9ki54 neisseria g
18	271	5.5	912	2	Q9KK96	Q9kk96 neisseria g
19	269.5	5.5	693	2	Q56644	Q56644 vibrio chol
20	269.5	5.5	915	2	Q9KK97	Q9kk97 neisseria g
21	268.5	5.5	698	16	Q9KM13	Q9km13 vibrio chol
22	266	5.4	912	2	Q9KI56	Q9ki56 neisseria g
23	265	5.4	791	16	Q9JYA8	Q9jya8 neisseria m
24	265	5.4	910	16	Q53348	Q53348 neisseria m
25	264	5.4	791	2	P72084	P72084 neisseria m
26	262.5	5.4	688	2	Q93SS7	Q93ss7 plesiomonas
27	261.5	5.3	791	2	P72085	P72085 neisseria m
28	258	5.3	883	2	O53136	O53136 neisseria m
29	255.5	5.2	915	2	Q9JP70	Q9jpp0 neisseria m
30	254.5	5.2	915	16	Q9KOU9	Q9kou9 neisseria m
31	249	5.1	940	2	O51187	O51187 neisseria m
32	248.5	5.1	742	16	Q9CN63	Q9cnc63 pasteurrella
33	247	5.0	792	2	Q9RGS9	Q9rgs9 neisseria m
34	246	5.0	784	16	O9CMS1	O9cms1 pasteurrella
35	246	5.0	792	2	O51104	O51104 neisseria m
36	244.5	5.0	716	2	Q9AJS1	Q9ajs1 vibrio angu
37	242.5	5.0	743	16	Q9ZN43	Q9zn43 rhizobium m
38	239.5	4.9	764	2	O68881	O68881 pseudomonas
39	239.5	4.9	764	16	Q9HV88	Q9hv88 pseudomonas
40	232.5	4.8	943	2	O50952	O50952 neisseria g
41	230.5	4.7	912	2	O48037	O48037 haemophilus
42	229	4.7	708	2	Q9XBV1	Q9xbv1 porphyromon
43	228	4.7	911	2	Q48044	Q48044 haemophilus
44	227.5	4.6	867	2	Q9KIB1	Q9kib1 porphyromon
45	227.5	4.6	943	2	O87343	O87343 neisseria m

#### ALIGNMENTS

RESULT 1

Q9JR50  
ID Q9JR50 PRELIMINARY; PRT; 921 AA.  
AC Q9JR50  
DT 01-OCT-2000 (Tremblrel. 15, Created)  
DT 01-OCT-2000 (Tremblrel. 15, Last sequence update)  
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)  
DE PUTATIVE TONB-DEPENDENT OUTER MEMBRANE RECEPTOR (TONB-DEPENDENT RECEPTOR).  
GN TDFH OR NMB1497.  
OS Neisseria meningitidis, and  
OS Neisseria meningitidis (serogroup B).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=487, 491;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=IR1074;  
RA Turner P.C., Thomas C.E., Stojiljkovic I., Elkins C., Kizel G.,  
RA Ala Aldeen D.A., Sparling F.P.;  
RT "Identification and Characterization of Putative TonB-dependent Outer Membrane Proteins among Pathogenic Neisseria".  
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.  
RN [2]  
RP SEQUENCE FROM N.A.  
RC STRAIN=MC58 / SEROGROUP B;  
MEDLINE=20175755; PubMed=10710307;  
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,  
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,  
RA Nelson W.C., Gwinn M.L., DeBoy R., Peterson J.D., Hickey E.K.,  
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,  
RA Mason T., Clecchi A., Parksey D.S., Blair E., Cittiore H., Clark E.B.,  
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J.,  
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,  
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;  
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58".  
RL Science 287:1809-1815(2000).  
DR EMBL; AF227418; AAF73907.1; -.  
DR EMBL; AF002499; AAF41853.1; -.  
DR

DR TIGR; NMB1497; -;  
DR InterPro: IPR000531; TonB\_boxC.  
DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; UNKNOWN\_1.  
KW Receptor; Complete proteome.  
SQ SEQUENCE 921 AA; 104222 MW; 67985B75EDF8819D CRC64;

Query Match 99.8%; Score 4883; DB 16; Length 921;  
Best Local Similarity 99.8%; Pred. No. 4.6e-259;  
Matches 919; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

QY 1 MRSSFLRKPICFYLMGVTLHYHYAEADAGRAGSEAOIQVLEDVHVAKRVPKDKKVFDTA 60  
DB 1 MRSSFLRKPICFYLMGVTLHYHYAEADAGRAGSEAOIQVLEDVHVAKRVPKDKKVFDTA 60

QY 61 RAVSTRQDIFKSSLENLDNIVRSIPGFTQDKSSGIVSLNIRGDSGFRVNTWVDGITQT 120  
DB 61 RAVSTRQDIFKSSLENLDNIVRSIPGFTQDKSSGIVSLNIRGDSGFRVNTWVDGITQT 120

QY 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGFSGSAGINSLAGSANLRTILGVDDVV 180  
DB 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGFSGSAGINSLAGSANLRTILGVDDVV 180

QY 181 QGNNTYGLLLKGLTGNTSTKGNMAAIGARKWLESGASVGLYGHSSRSVAQNYRVGGGG 240  
DB 181 QGNNTYGLLLKGLTGNTSTKGNMAAIGARKWLESGASVGLYGHSSRSVAQNYRVGGGG 240

QY 241 QHIGNFGAEXLERRKQRYFVOEGALKENSQKWERDLQROQWKYKPKYNNQELQYI 300  
DB 241 QHIGNFGAEXLERRKQRYFVOEGALKENSQKWERDLQROQWKYKPKYNNQELQYI 300

QY 301 EGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNFKIGSRK 360  
DB 301 EGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNFKIGSRK 360

QY 361 IINRNYQFNGLSLNPYTNLNTAAYNSGRQKYPKSGKFTGWGLLKDFEYNNAKIIDLN 420  
DB 361 IINRNYQFNGLSLNPYTNLNTAAYNSGRQKYPKSGKFTGWGLLKDFEYNNAKIIDLN 420

QY 421 NTATFRLPRETELQTTIGFNYFHEYKGNRPPEELGLFFPDQDNGLSYLGFRKGDGK 480  
DB 421 NTATFRLPRETELQTTIGFNYFHEYKGNRPPEELGLFFPDQDNGLSYLGFRKGDGK 480

QY 481 LLPQKSTIVQAGSQYFNTFYDAAALKDIYRLNYSNTNTVYRFGGEYTYGYSDDDEFKR 540  
DB 481 LLPQKSTIVQAGSQYFNTFYDAAALKDIYRLNYSNTNTVYRFGGEYTYGYSDDDEFKR 540

QY 541 AFGENSPYKHKCNRSQGIYEPVLKYYKKRANNSHVSISADFGDYFMPFASYSRTHRP 600  
DB 541 AFGENSPYKHKCNRSQGIYEPVLKYYKKRANNSHVSISADFGDYFMPFASYSRTHRP 600

QY 601 NIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDPLGLKLVGYRSRIDNYI 660  
DB 601 NIQEMYFSQIGDSGVHTALKPERANTWQFGFNTYKGLLKQDDPLGLKLVGYRSRIDNYI 660

QY 661 HNVTGKWNLDNGDTPSWSSSTGLAYTTOHRNFKDKVHKHGFELNLDYGRFFTNLSYAY 720  
DB 661 HNVTGKWNLDNGDTPSWSSSTGLAYTTOHRNFKDKVHKHGFELNLDYGRFFTNLSYAY 720

QY 721 OKSQPTNFDASPSNNASKEDQKQYGLSRVSALPRDYGRLVGTWRGNKLTILGGA 780  
DB 721 OKSQPTNFDASPSNNASKEDQKQYGLSRVSALPRDYGRLVGTWRGNKLTILGGA 780

QY 781 MRYFGKSIRATAERYIDGTNGGNTSNFRQLGKRSIKQTETLARQPLIFDFYAAEYKKN 840  
DB 781 MRYFGKSIRATAERYIDGTNGGNTSNFRQLGKRSIKQTETLARQPLIFDFYAAEYKKN 840

QY 841 LIFRAEVKNLFDTRYIDPLDAGNDAATQRYSSDPDKDKEDVTCNADKTLGNGKYGGTS 900  
DB 841 LIFRAEVKNLFDTRYIDPLDAGNDAATQRYSSDPDKDKEDVTCNADKTLGNGKYGGTS 900

QY 901 KSVLTNFAFGRTFLMTMSYKF 921  
DB 901 KSVLTNFAFGRTFLMTMSYKF 921

Db 901 KSVLTNFAFGRTFLMTMSYKF 921

RESULT 2  
Q9JTN4 PRELIMINARY; PRT; 922 AA.

ID Q9JTN4  
AC Q9JTN4;  
DT 01-OCT-2000 (T-EMBLrel. 15, Created)  
DT 01-OCT-2000 (T-EMBLrel. 15, Last sequence update)  
DT 01-DEC-2001 (T-EMBLrel. 19, Last annotation update)  
DE PUTATIVE OUTER MEMBRANE SUBSTRATE BINDING PROTEIN.  
GN NMA1700.  
OS Neisseria meningitidis (serogroup A).  
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.  
OX NCBI\_TaxID=65699;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;  
RX MEDLINE=20222556; PubMed=10761919;  
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C., Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T., Davies R.M., Davis P., Devlin K., Mungall K., Quail M.A., Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J., Whitehead S., Spratt B.G., Barrell B.G.;  
RA "Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491";  
RL Nature 404:502-506(2000).  
DR EMBL; AL162756; CAB84928.1; -;  
DR InterPro: IPR000531; TonB\_boxC.  
DR PROSITE: PS01156; TONB\_DEPENDENT\_REC\_2; UNKNOWN\_1.  
KW Complete proteome.  
SQ SEQUENCE 922 AA; 104290 MW; B45E683985DE20B6 CRC64;

Query Match 97.3%; Score 4763.5; DB 16; Length 922;  
Best Local Similarity 97.5%; Pred. No. 1.6e-252;  
Matches 899; Conservative 9; Mismatches 13; Indels 1; Gaps 1;

QY 1 MRSSFLRKPICFYLMGVTLHYHYAEADAGRAGSEAOIQVLEDVHVAKRVPKDKKVFDTA 60  
DB 1 MRSSFLRKPICFYLMGVTLHYHYAEADAGRAGSEAOIQVLEDVHVAKRVPKDKKVFDTA 60

QY 61 RAVSTRQDIFKSSLENLDNIVRSIPGFTQDKSSGIVSLNIRGDSGFRVNTWVDGITQT 120  
DB 61 RAVSTRQDIFKSSLENLDNIVRSIPGFTQDKSSGIVSLNIRGDSGFRVNTWVDGITQT 120

QY 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGFSGSAGINSLAGSANLRTILGVDDVV 180  
DB 121 FYSTSDAGRAGSSQFGASVDSNFIAGLDVVKGFSGSAGINSLAGSANLRTILGVDDVV 180

QY 181 QGNNTYGLLLKGLTGNTSTKGNMAAIGARKWLESGASVGLYGHSSRSVAQNYRVGGGG 240  
DB 181 QGNNTYGLLLKGLTGNTSTKGNMAAIGARKWLESGASVGLYGHSSRSVAQNYRVGGGG 240

QY 241 QHIGNFGAEXLERRKQRYFVOEGALKENSQKWERDLQROQWKYKPKYNNQELQYI 299  
DB 241 QHIGNFGAEXLERRKQRYFVOEGALKENSQKWERDLQROQWKYKPKYNNQELQYI 300

QY 300 IEGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNFKIGSR 359  
DB 301 IEGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNFKIGSR 360

QY 360 KIINRNYQFNGLSLNPYTNLNTAAYNSGRQKYPKSGKFTGWGLLKDFEYNNAKIIDL 419  
DB 361 KIINRNYQFNGLSLNPYTNLNTAAYNSGRQKYPKSGKFTGWGLLKDFEYNNAKIIDL 420

QY 420 NNTATFRLPRETELQTTIGFNYFHEYKGNRPPEELGLFFPDQDNGLSYLGFRKGDGK 479  
DB 421 NNTATFRLPRETELQTTIGFNYFHEYKGNRPPEELGLFFPDQDNGLSYLGFRKGDGK 480

QY 480 GLLPQKSTIVQAGSQYFNTFYDAAALKDIYRLNYSNTNTVYRFGGEYTYGYSDDDEFK 539  
DB 480 GLLPQKSTIVQAGSQYFNTFYDAAALKDIYRLNYSNTNTVYRFGGEYTYGYSDDDEFK 539





```

Db 451 --KEGFNPSVQLSAQVTPWLQPPICGYSKSMRAPNIQEMFFNSGGASMPNPKPERAE 507
QY 626 TWQFGFNTYKGLLKQDDTLGLKLVGRSIRDNIYHN-----VYQKWD 669
Db 508 TWAQGFNIDRLDLEQDALRFAALAYRSIQNYIYSELYLVCSSGGRKCSMAEIVGNDWE 567
QY 670 LMGDIPSWSSSTGLAYTIQHRNEFKVHKHGFEELELNDYGRFTNLNLSYAKSTQPTNF 729
Db 568 --GISDEYSDNIYV-----NSASDVIAKGFLEMDYDAGFAFGRUSFSQOQTDQPTSI 620
QY 730 SDASESPNNASKEDQLKQGLRVSALPRDYGRLVGTWRWLNKLTGLGAMRYFGKSIR 789
Db 621 ASTY-----FGAGDMTELPKRYMTLDGVRFFDNALTLGLTIKYTKARR 665
QY 790 ATABERYIDTNGTNSFNQLGKRSKTQETLARQPLIFDFAAYEPKKNLIFRAEVKN 849
Db 666 LSPDFEODEHTGA-----IIKQ--DLQPIITIIDLYGYEYRNLTKLSVQN 711
QY 850 LFDRIYDPLDAGNDAATQRYSSFPDKDDEVTGNADKTLGCKYGGTSKSVLTNFA 909
Db 712 LMNDYSEALNKLN-----MMPGLGDETHPANS-----AR 741
QY 910 GRTEFLITMSYKF 921
Db 742 GRTWIFGGDIRF 753

RESULT 6
Q92NX4
ID Q92NX4 PRELIMINARY; PRT; 885 AA.
AC Q92NX4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PUTATIVE IRON/HEME TRANSPORT PROTEIN.
GN SMC04205.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium;
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Leilaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA Ramsperger U., Surzycki R., Thebaut P., Vandenbol M.,
RA Vorhoeffer F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RL Science 293:668-672(2001).
DR ENBL; AL591789; CAC46612.1; -.
KW Complete proteome.
SQ SEQUENCE 885 AA; 95389 MW; BEE36641060DFBF4 CRC64;

```

Query Match 9.6%; Score 468.5; DB 16; Length 885;  
 Best Local Similarity 22.8%; Pred. No. 1.2e-17;  
 Matches 228; Conservative 122; Mismatches 337; Indels 313; Gaps 42;

```

QY 14 LMGVTLYHSY-----AEDAGRAGSEAQIQVLEDDVHVAKRVPKDK----- 55
Db 107 LLDGTGVHYQFTANRAAVICPAGDAGPGSEEGATVLRIVVTGKTCRNANSAGFGCTP 166
QY 56 --VFTDARAYS-TRQDIFKS--SENLDNIYRSIPGFTQDQKSSGI-VSLNIRGDSGFR 109
Db 167 DWYEEPASVSVSDAVQSKAARNANDVLDVSGVTSNRSEAGNPGIAINVRGLQDNR 226

```

```

QY 110 VNTWVDGIITTFYSTSTDAGRAGSSQFGASVDNSFNITAGLDVYKSGSGSAGINSLAGSA 169
Db 227 VTTMIDGARQDFQ-----RAGHGASQRYVVDYTAFLRSVEVEKGVAGVGGASLGGAV 279
QY 170 NLRTLGVDVVOGNTYGLLLKGLTGNSTKGNAMAIGARKWLESGASGVLYGHRRS 229
Db 280 NFRVTADDDIITPDGRGVNELNAETGYNAYFNG-SLIGARFSE---DSVLGGISKR 335
QY 230 VAONYRVGGGQHGIGNFAGBYLERRKORYFVQEGALKFNSDGKWERDLORQOQKYPKY 289
Db 336 VG-DYDFGQNG-----KSPLLDLAVTTAVDDSDFLFSR----- 366
QY 290 NYNNQELQKY-----IEG-----HDKSWRENLAPOYDITPI-----DPSCLK 326
Db 367 -----LETEGTLKVEGSPSDFTFDLSWLN-----DSEATQGGVFGDLRDPQNYL 415
QY 327 QQSAGNLFKLEYGCVFNKYTAQFRDLNFKIGSRKIIINRNYQFNGLSNPYT-NNLNTAA 385
Db 416 NNTVSSSEFWDDPD-----SELIDKGLRWYNRVND-----ELROYTPPLPIYA 460
QY 386 YNSGRQKYPKSGFTGWGLKDFETYNNAKILDJNNTATFRLPRETELQTTLGFNY---F 442
Db 461 MTS-----FGG-----SLDNTSRF-----ETAL-GDLSLNYGGEA 489
QY 443 HNEYGKNRFPPEELGLFDFDPDQDNGL---YSYLGFRFGDKGLLPQKSTIVQPGASQYFNT 499
Db 490 YSDNGKTTTPPLV-----DQGFDEAYGKG-----LNPVGRSRMTS 526
QY 500 FYFD-----AALKKIYRLNYSNTVVG-----YRFGGEBYGYGSDDEFFKRAFG 543
Db 527 AFLNATLEHDDWLEVGAGLGRDYRLKAGFTEVGGRKPRYIVPGVCGYFYDDGEC---AYY 584
QY 544 ENSPTYKKHCNRSGIYEPVLKYGKKRANN---HSVSIASADFGDYPMPFASYSRTHRMP 600
Db 585 DEDPVY-----GGGEAVLERVDIDKSGCALLPSARIAVMPEGIPQFVYTAHTYRPP 636
QY 601 NIQEMYSQIGDSGVHTA-----LKPERRANTWQGFNTYKGLLKQDDTLGLKL 649
Db 637 SVMKAL-----TSGGHPGDAIATYIPNPKRPERGTWELGINIARDGLFTAGDSLRLKT 691
QY 650 VGYRSRIDNYI--HNVYCKWMDLNGDIPSWSVSTGLAYTIQHRNFKDKVHKHGFEELELNY 707
Db 692 YPFDRTIQTITLNGYFAFDKN-----LFQHVNLDDGDTTMNGVEIEASY 737
QY 708 DYGRFHTNLSYAVOKSTQPTNFS-----DASESPNNASKEDQLKQGLYSLRVSALPRDYGR 763
Db 738 DMGSAYVGASYTLKTDYADTYSYSGPTASGTPLAASGNTPPV-----SVLFVPPENKFT 792
QY 764 LEVGTWMLGNKLTILGGAMRYFGKSIRATAEERYIDGTNGG--NTSNPRLGKRSIKQET 821
Db 793 LDAGIRLFERKLVLGGRATY-----VSDSKPTVGQLAGLFTAGYK----- 833
QY 822 LARQPLIFDEYAAVEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFPDKDKDE 881
Db 834 -----VEDIYGSYFSDSAKLRLAINNVDEQYAPALGA-----FYYP----- 871
QY 882 DVTGNADKTLGCKYGTSKSVLTNFAGRFTFLITMSYKF 921
Db 872 -----APGRATATVSLNFKF 885

RESULT 7
Q9X5P4
ID Q9X5P4 PRELIMINARY; PRT; 883 AA.
AC Q9X5P4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HEME RECEPTOR HASR.
GN HASR.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.

```

```
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20121752; PubMed=10658665;
RA Ochster U.A., Johnson Z., Vasil M.L.;
RT "Genetics and regulation of two distinct haem-uptake systems, phu and
RL Microbiology 146:185-198(2000).";
DR EMBL: AF127223; AAD31013.1; -
DR InterPro: IPR005531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
KW Receptor.
SQ SEQUENCE 883 AA; 96942 MW; A29B07A3A6C31DCF CRC64;

Query Match 8.9%; Score 435.5; DB 2; Length 883;
Best Local Similarity 23.0%; Pred. No. 7.9e-16;
Matches 218; Conservative 117; Mismatches 375; Indels 239; Gaps 39;

QY 30 RAGSEAIQVLEVDVHKAKRVPKDKKVFDDARAVST--RQDIFKS-SENLDNIVRSIPGA 86
Db 117 RVSQDLDVQMSPSV-ISAAR--PDDWVYQTPHSVSVIGREIERNPRHAADMLEETPGV 173

QY 87 FTQODKSSGIVSLNIRGDSGFRVNTWVDGTTOTFTYSTSTADAGAGSSQFGA-SVDSNF 145
Db 174 YSSVSQDDPGLSVNIRGIQDYGVRNMSVDGMQRNYQ-----QSGHQQRNGTLYVDPEL 226

QY 146 IAGLDVVKVSGSAGINSIAGSANLRTLGVDVVOGNNYTG---LLLKGLTG-TNSTK- 200
Db 227 LSEVVIDKGASSAMGAGVIGGIANFRTLEARDLVPRGKQVGRVRLTSGLGDANGTHF 286

QY 201 -GNAMAAIGARKWLESASVGLYGHSSRVAQNYRVGGGQHGQIFGAELERRKQRYF 259
Db 287 IGSAFAAIGTEW-----DMLVAASERHLG-DYDPGTKGS-IGE----- 323

QY 260 VQEGALKFNDSGKWERDLQROQWKYP--YKNYNQELQKYIEGHDKSWRENLAPQYDI 317
Db 324 LRTGAW-FNPEAG-----QVRKHSVAVSGY-----VMSRLAKLGA 360

QY 318 TPIDPSSLKQOAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNGLSL-NP 376
Db 361 LPDQO---RLQFSYLTQVSYDDA-NMLTENQALMEKLGSSDVRAQNFADYGYAPDNP 416

QY 377 YTNLNLTAAY--NSGRQKYPKSGKFTGWLLKDFEYNNNAKIIDLNTATFRLPRETELQ 434
Db 417 LVDFKAKLYVDNRNQRTLQRTGYSITYQTDYVG---AAQNTSTFALDLDLTLR 472

QY 435 TTIGFNVFHNKYNR-----FPEELGLFFDGPDPDNGLYSLGRFRKDGKGLLP 483
Db 473 ANYGLEFFYDKVRPSSQPRASTSAVGFPAAEGM---TPKGRALGSLFARLDYD----- 524

QY 484 QKSTIVQAGSOYFNTFYDAALKDIYRL-----NYSNTVTYGRFGGEYTYGSDDE 537
Db 525 -----YDDMLNLNAGLRYDRYLRGTGTGFNARTFLLGTTRQTDMPLOYAVD-- 570

QY 538 FKRAFGENSPTYYKHKHCRSGGIVEPVLKKGKRRANNHVSISADFG-DYFMPFASYSRT 596
Db 571 --REGRFSPTF-----GLSVKPGVDWLQLFATYKG 600

QY 597 HRPNPTQMYFSIGDSG-----VHTALKPERANTWQGFNTYKGLLKQDDTLGLKLV 650
Db 601 WRPPAYTESLITGRPHGGGAENMYPNPLSPERSKAWGVFNVLKENLWFSDDLGLKVA 660

QY 651 GYBSRIDNVIHNYGKWDLNGDIPSWSVSTGLAYTIOHRNFKDKVHKHGFLELNDYDG 710
Db 661 YFTRVDDDFIFMGMGQ-----PPGYGMAGIGNSAYVNNL--DSTRFRGEVYQLDYDAG 712

QY 711 RFTNLISYAYQKS-----TQPTNFSDASESPPNASKEDQLKQGYGLSRV-- 754
Db 713 LAYGQLSYTHMIGNSDFCSKTAWLGGVTVTKGSGRRPPVIMDRPDEQANATHCSAVLG 772

QY 755 --SALPRDYGRLEVGRVRLGNKLTLLGAMRYFGKSTRATAERYIDGTNGGNTSNFRQLG 812
```

```
Db 773 SAEHMPMDRG-----SLTLG--MRFFORRLDVGARARYSEGYSVAGGATVSOAG 819
QY 813 KRSIKQTEILARQPLIFDFAAYEPKKNLIFRAEVKNLFDKRRYIDPLDAGNDAATQRYYS 872
Db 820 VYPADWKY-----TVYDLYGSYRVSDLEFLRLAMENVTDVRLVPLG----- 862
QY 873 SFDPKDKDEDTVCNADKTLNCGKYGKTSKSVLTNFARGRTFLITMSYKF 921
Db 863 -----DV-----LAFTLGRGRTLOGTLEQYF 883

RESULT 8
Q9HYJ7 PRELIMINARY; PRT; 891 AA.
AC Q9HYJ7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DE HEME ACQUISITION PROTEIN HASR.
GN HASR OR PA3408.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltry L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL: AB004762; AAG06796.1; -
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 891 AA; 97864 MW; 126D6F41E7D66979 CRC64;
```

Query Match 8.9%; Score 435.5; DB 16; Length 891;  
Best Local Similarity 23.0%; Pred. No. 8e-16;  
Matches 218; Conservative 117; Mismatches 375; Indels 239; Gaps 39;

```
QY 30 RAGSEAIQVLEVDVHKAKRVPKDKKVFDDARAVST--RQDIFKS-SENLDNIVRSIPGA 86
Db 125 RVSQDLDVQMSPSV-ISAAR--PDDWVYQTPHSVSVIGREIERNPRHAADMLEETPGV 181

QY 87 FTQODKSSGIVSLNIRGDSGFRVNTWVDGTTOTFTYSTSTADAGAGSSQFGA-SVDSNF 145
Db 182 YSSVSQDDPGLSVNIRGIQDYGVRNMSVDGMQRNYQ-----QSGHQQRNGTLYVDPEL 234

QY 146 IAGLDVVKVSGSAGINSIAGSANLRTLGVDVVOGNNYTG---LLLKGLTG-TNSTK- 200
Db 235 LSEVVIDKGASSAMGAGVIGGIANFRTLEARDLVPRGKQVGRVRLTSGLGDANGTHF 294

QY 201 -GNAMAAIGARKWLESASVGLYGHSSRVAQNYRVGGGQHGQIFGAELERRKQRYF 259
Db 295 IGSAFAAIGTEW-----DMLVAASERHLG-DYDPGTKGS-IGE----- 331

QY 260 VQEGALKFNDSGKWERDLQROQWKYP--YKNYNQELQKYIEGHDKSWRENLAPQYDI 317
Db 332 LRTGAW-FNPEAG-----QVRKHSVAVSGY-----VMSRLAKLGA 368

QY 318 TPIDPSSLKQOAGNLFKLEYDGVFNKYTAQFRDLNTKIGSRKIINRNYQFNGLSL-NP 376
Db 369 LPDQO---RLQFSYLTQVSYDDA-NMLTENQALMEKLGSSDVRAQNFADYGYAPDNP 424
```



```
QY 377 YTNLNTAA--NSGRQKPKGSKFTGWLKQFETVNNAKIILDLNNTATRLPRETELQ 434
Db 425 LVDFKAKLYVDNRNROQTQRTGTTGYSITYQTDYTG----AQAOQNTSTFALDLSLTLR 480
QY 435 TTLGFNFHNEYGNR-----FPEELGLFFDGPDDNGLSYLGFRGDKGLLP 483
Db 481 ANYGLEFFYDKVRDPSQPRASVSAGVFPAAEGM---TPKGDRAIGSLFARLYD----- 532
QY 484 OKSTIVOPAGSOYFNTFYFDAALKKDIYRL-----NYSTNTVYRGFGGYTYGYSDDDE 537
Db 533 -----YDDLNLNAGRLRYDRYLRGDTGFNARTFILTTRQTDMPLOYAVD-- 578
QY 538 FKRAFGENSPTYKKHCNRSCGIYEPVLKKYKKRANHHSISADFG-DYFMPFASYSRT 596
Db 579 --REEGFRSPTE-----GLSVKPGVDWLQLFATYKKG 608
QY 597 HRMPNIQEMTFSGDGS-----VHTALKPERANTWQFGFNTYKGLLQDDDTGLGLV 650
Db 609 WRPPAVTESLITGRPHGGGAENMYPNFELSPERSKAWEVGFNVLKENLWFSDDDELGLKVA 668
QY 651 GYRSRIDNYIHNVYKWMWDLNDIPSWSTGLAYTTOHRNFKDKVHKHGFELNELNYDYG 710
Db 669 YFDRVDDFIFMGGMQ-----PCGYGMAGIGNSAYVNNL-DSTRFRGVVEYQLDYDAG 720
QY 711 RFTNLISYAYOKS-----TQPTNFSDAESPNNASKEDQLKOGYGLSRV-- 754
Db 721 LAYQLSYTHMIGNSDFCSKTAMUGVTVQVKGSGRRPPVIDMRPEQANAATHCSAVLG 780
QY 755 --SALPRDYGRLEVTGRLGNKLTGLGAMRYFGKISIRATAEERYIDGTNGNTSNFRLQ 812
Db 781 SAHWPMDRG-----SLTLG--MRFDFRLDVGARARYSEGYVAGGATVSOAG 827
QY 813 KRSTKQETLARQPLIFDFAAEPEKKNLIFRAEVKNLFDORRYIDPLDAGNDAATQRYYS 872
Db 828 VIPADWKEY-----TVYDLYGSYRVSDELTLRLAMENVTDRAYLVPLG----- 870
QY 873 SFDPKDKDEDVCNADKTLGCKYGGTSKVLNTNFARGRTFLITMSYKF 921
Db 871 -----DV-----LAFTLGRGRTLQGTLEYQF 891

RESULT 9
Q9HUR6
ID Q9HUR6 PRELIMINARY; PRT; 989 AA.
AC Q9HUR6
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, Last sequence update)
DT 01-OCT-2001 (Tremblrel. 18, Last annotation update)
DE HYPOTHETICAL PROTEIN PA4897.
GN PA4897.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004902; AAC08282.1; -
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 989 AA; 108303 MW; 99D332FDBF0806F CRC64;
```

```
Query Match 8.8%; Score 428.5; DB 16; Length 989;
Best Local Similarity 23.0%; Pred. No. 2.2e-15;
Matches 229; Conservative 132; Mismatches 375; Indels 261; Gaps 42;

QY 30 RASEAQIOVLEDDVHVAKRVPKDKVFTDARAVSTQDIFKSSENLD-----NIVR 81
Db 149 RMSGEAPADLSPPVVVSAEELADPOKETYT-----APRSSVYLSEEDIRFORVSVGDLQ 203
QY 82 SIPCAFTQODKSSGIVSLNRGSGFGRVNTVDTGTQTFYSTDAGRAGGSGQFASV 141
Db 204 GIFCVQVDSRNGALDVNIRIGQSRVAVRVDCAEQ-----ALDVYRGVAGTQORSYI 258
QY 142 DSNFIAGLDVYVFGSFGSAGINSIAGSANLRLTGVDDVQGNNTYGLLLKGLTGTNSTKG 201
Db 259 DPDLVSVTVTKGSPTRSGAI---GGSVEMRTIGVKDILVDGKDLGVFRFTGDVNN--G 312
QY 202 NAMAAGIARKWLESGASY---GVLYGHSPR--SVAQNYRVGGGGHIGNFGAEYLERR 254
Db 313 VAPQHRASAKTENLSSVPHDDRGSLFGSAKSGSAAPAYR---NEHL-DLVAAYAQNR 367
QY 255 KQYFVQEGALKFNSDSGKWERDLQROQWKYPYKYNNOE---LQKYIEGHD----- 304
Db 368 QGNYF-----SGKKQD-----RYRYNRYGRESSVAKYVYNAGEEVLNSESSE 410
QY 305 -----KSWRENLAPOYDIT-----PIDPSLKQOOSAGNLFKLEYDGV-FNKY 345
Db 411 TESYLLKATWR--LADSETLDLGYRRYDGRGTGEIMPSDIFRGTAGIYQVPLSEVKIDTY 468
QY 346 TAQFRLNTKIGSKIIINRYQFNYSGLSLAPYTNLNLTAAYNSGRQYKPGSKKTGWLGL 405
Db 469 TARYRILPEN-----NPLVDLSTGLWMTAKSMDLTSVLAPRSQARYSRDNWT----- 516
QY 406 KDFETYNNAKI-LDLNNTATFRL-----PRETELOTTLGFN----- 440
Db 517 ----RQDNRRIIGDLNNVARETDFGDKLDLGGSFQVEDIQPKSVVTVLHDINANRTL 572
QY 441 --YFHNIEYGN-----RPPEELGLFFDG-----PDQONGLSYLGFRGDKGLLPKQSTI 488
Db 573 RDATRQYGLNGKLEFRPVERLTLWGGGRYSHFNKSGISASPRREDRDMRFI---TV 628
QY 489 VQPA--GSOYF---NTFYFDA--LKLDIYRLNYSNTVYRGFGGYTYGYGSDDEBKF 539
Db 629 SRPGYYSMMWFFPDQNGYTDATDPRLNNGIVTNTNPPFEGIFP-----DEFG 677
QY 540 RAFGENSPYKKHCNRSCGIYEPV-LKKYKKRANHNSVSIADSF---GDYFMPFASYR 595
Db 678 PA-----NVTVHPSRVTNVVTVGYNSYKKGSSRGGSFAPFGINFELAPDFTV-YASYTE 730
QY 596 THRPNTQEMYFSOIGSDGVHTA--LAPERANTWQFGFNTYKGLLQDDTLGLKLVGR 653
Db 731 GLRLPSLFE---TSQGTLOVEPGKDLKPERSRWEIGASALRDSLLADGDSAAIKLAYFN 787
QY 654 SRIDNYIHNVYKWMWDLNDIPSWVSTGLAYTTOHRNFKDKVHKHGFELNELNDYGRFF 713
Db 788 NTIKNYTRYD-----PGQGLMTFSTNT-----DSYRTSGLELQSHDAGRVF 831
QY 714 TNLISYAKSTQPTNFSDAESPNNASKEDQLKOGYGLSRVSAL-----PRDYGR 764
Db 832 ADLSATYLYLTETCDAAFAARLAGANRYQRTENTPCTPGSPMGSYTNTQNPRLATNL 891
QY 765 EVGTRWLGKNTLGGAMRYFGKSTRATAERYIDGTNGNTSNFRLQGRKRIKOTETLAR 824
Db 892 TAGLRFDDQALTLGGRMTY-TSGGTATADKPKWQV-----ATTPQIERYSVQ----- 937
QY 825 QPLIFDFAAEPEKKNLIFRAEVKNLFDORRYIDPLDAGNDAATQRYYSFDPKDKDEDVT 884
Db 938 ---LFDLFLYKLFKEHTLNLASLQNLTDRYLYDPL-----AQSFMP----- 975
QY 885 CNADKTLGCKYGGTSKSVLNTNFARGRTFLITMSYKF 921
Db 976 -----APGRTLVRGMQAKF 989
```

```
RESULT 10
Q9RHT4 PRELIMINARY; PRT; 916 AA.
AC Q9RHT4;
DT 01-MAY-2000 (TrEMBLrel. 13, Created)
DT 01-MAY-2000 (TrEMBLrel. 13, Last sequence update)
DT 01-JUN-2001 (TrEMBLrel. 17, Last annotation update)
DE HSA RECEPTOR PROTEIN.
GN HASR.
OS Pseudomonas fluorescens.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=294;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=NO.33;
RX MEDLINE=20069636; PubMed=10601212;
RA Idei A., Kawai E., Akatsuka H., Omori K.;
RT "Cloning and characterization of the Pseudomonas fluorescens ATP-
binding cassette exporter, HasDEF, for the heme acquisition protein
HasA.";
RL J. Bacteriol. 181:7545-7551(1999).
DR EMBL; AB023289; BAA88490.1;
DR InterPro; IPR001589; Actinin_act_bind.
DR Pfam; PF00593; TonB_boxC.
DR PROSITE; PS00019; ACTININ_1; UNKNOWN_1.
KW Receptor.
SQ SEQUENCE 916 AA; 101187 MW; CF392CAA8939EDC8 CRC64;

Query Match 8.7%; Score 426; DB 2; Length 916;
Best Local Similarity 22.5%; Pred. No. 2.7e-15;
Matches 226; Conservative 129; Mismatches 372; Indels 276; Gaps 45;

QY 22 YSYAEDAGRAGSEAQIQ--VLEDVHVAKRVPKDKK-----VFT 58
DB 87 YSAEGLGLRGVSAVPYRFGDGRQVLTTRVSADHDGALTGVGTTITGNTGKPSDWYD 146
QY 59 DAR--AVSTRQDIFK--SSELDNIVRSIPGAFQODKSSGIVSLNIRGDSGFRVNTMVD 115
DB 147 TPRSVAVISREQIDKRPRIADMLEETAGVYTAVNQRDPGLSVNIRGVQDYGRVNNID 206
QY 116 GITQTEYSTDAGRAGSSQFGAS--VDSNFIAGLDVVKSGSPGSAGINSLAGSANLRT 173
DB 207 GMQNF-----NVNGHQORNGTMLIDPEFISIEIDKSGSGQGGAVALGGIASFKT 258
QY 174 LGVDVVYQGNNTYGLLLKGLTGT----NST--KGNMAAICARKWLESGASGVLYGHSR 227
DB 259 LEASEFLADKDYGRLRAGSGIGELNGTYFNGSGVFAFGD---ERG---DVLLGYSE 311
QY 228 RSVANQYRVG--GGQHGIGNGAELYERKORYFVQEGALK--FNSDSGKWERDLRQOWK 284
DB 312 RHFG--DYRAGTHNDKQLGT----HLRADS-----QPAAFDDDLNLSVEGDTGTSVRSQIVK 362
QY 285 Y---KPKYNNQBELQ--KYIEGHDKS---WR-----ENLAPQYDITPIDSSLKQOQAGNL 333
DB 363 FGLNLP-----NDQRVQLSYLESDDSDNDAYATPDNQSYVQ--RVSKNNLNKAKNVGLD 416
QY 334 FKLEYDGVFNKYTAQFRDLNKTGSRKIINRNYQFNVYGLSLNPTNLNLTAAVNSGRQKY 393
DB 417 YRYTPDNPMIDFRAKVYVYVTQMDRENAPN-----AATLTSGNVYVAGVTDHFQT- 465
QY 394 PKGSKFTGWLGLDDFEYNNAKILDLNNTATFRLPRETELQTLTGFNYFNEY--GKNRF 451
DB 466 -----DTWQVQGD-----NTSRFDFDALGHVSRNYGVVEYQDKFKPSTNKV 506
QY 452 PEELGLFFDQPDQDNGLYSL--GRFKGDKGLLPQKSTIVQAGSQVFNTFYFAALKKOI 510
DB 507 -EATNL-----SGLLPVEVGGTGGKRTI---ASLFNNLQYEGDWLTLEAGLYDR 554
```

```
QY 511 YRLNYSNTVGYRFGGEYTYGSDS-----EFKRAFGENSPYTKKHCNRSCTGIEPVLK 565
DB 555 YRLEGVTGMLTYRRDRFYSSVTGAKRVEEVFDIDREEGQFSPTF-----GIGI-KPCL- 606
QY 566 KYCKKRANHSVSISADFGDYFMPFASYSRTHRPNTQEMVFSQIGSGVHTA----- 618
DB 607 -----DWMQLYARWGKWRPPAVTETFTMTGRPHGSSSERVFPNPF 647
QY 619 LKPERANTWQGFNTYKGLLKODDTLGLKLVGYSRIDNYI---HNVYCKWMLNDGIDP 675
DB 648 LKPEESRDWEVGNVFKEGLLFGDRLGMKVAYEDTRIENFSFLNHSV-----SLP 698
QY 676 SWSSTSLAYTIQHRNFKDKVHKHGFELNLDYDGRFFTLNSYA----- 719
DB 699 E-TSVGGLGTGMAYVNTNTRFRGVEYQLNYDMGRAYANLSYTHMTGSGNEFCSKNYMG 757
QY 720 -----YOKSTOP-----TNFSDASESNNAKEDQLKQGYGLSRVSLALPR 759
DB 758 GAKKNGPSTTRYERYTRPNTGTIGLRPVTVTVTEVLDDDAAN-NKE---SCGRIMGNATYMPA 813
QY 760 DYGRLEVGTEWLNKLTGLGAMRYFGKSIRATAEERYIDGTNGGNTSNFRLGKRSIKQT 819
DB 814 DRGSLTLGARFLKRLDMGVVRVY-----SSNGENLDSQGYDFMDQ- 855
QY 820 ETLARQPLIFDFYAAAYEPKKNLIFRAEVKNLFDRIYIDPL-DAGNDAATQRYSSFDPKD 878
DB 856 -ALWPQYTLVDLYASYWMTDLNLTALALENATDEAYFVAMGDANN----- 899
QY 879 KDEDVTCNADKTLGNGKYGTSKSVLTNFARGRTFLITMSYKF 921
DB 900 -----LSLARGRTLGMLEYKF 916

RESULT 11
P72473 PRELIMINARY; PRT; 899 AA.
AC P72473;
DT 01-FEB-1997 (TrEMBLrel. 02, Created)
DT 01-FEB-1997 (TrEMBLrel. 02, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HASR PROTEIN.
GN HASR.
OS Serratia marcescens.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Serratia.
OX NCBI_TaxID=615;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SM365;
RX MEDLINE=97315228; PubMed=9171402;
RA Ghigo J.M., Letoffe S., Wandersman C.;
RT "A new type of hemophore-dependent heme acquisition system of Serratia
marcescens reconstituted in Escherichia coll.";
RL J. Bacteriol. 179:3572-3579(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=SM365;
RA Binet R.;
RL Submitted (OCT-1996) to the EMBL/GenBank/DBJ databases.
DR EMBL; Y08983; CAA70172.1;
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
SQ SEQUENCE 899 AA; 98281 MW; 3A13AE4CFCA3911D CRC64;

Query Match 8.4%; Score 412; DB 2; Length 899;
Best Local Similarity 23.4%; Pred. No. 1.6e-14;
Matches 221; Conservative 129; Mismatches 323; Indels 270; Gaps 48;

QY 56 VFTDARAVSTRQDIFKSSSENLDN-----IVRSIPGAFQODKSSGIVSLNIRGDSGF 107
DB 150 VYDEPRSVSV-----ISREQMDNRPARHADILEQTTTGAYSSVSQODPALSVNIRGIQDY 204
```

QY 108 GRVNTWVDGTTQTFYSTSTSDAGRAGGSGQFCASVDSNFIAGLDVVKGSFSGSAGINSIAG 167  
Db 205 GRVNNIDGMRFQKSG--HGQRGTMY-----IDSELLSGVTIDKGTGGMGAGTGG 258  
QY 168 SANLRTLGVDVQGNNTYIGLLGLGTGNTSK--GNAMAAIGARKWLESAGSVGLVGH 225  
Db 259 IATNTVYSADFAPRELGLKLGKLGKLGKLGKLGKLGKLGKLGKLGKLGKLGKLGKLGK 311  
QY 226 SRRSAQNYRVGGGQGHIGNFGABYLERRKORYFVQEGALKFNDSKQWDLRQOQWKY 285  
Db 312 SERHLG-DYMPGNKGD-IGN-----IRINDTGNDRYAESIKNNK 350  
QY 286 KPYKNYN-----NOELO-KYIEGHDKSWHENLAPQYDITPDPSSLKQOSA 330  
Db 351 IPDTHYRMHSLAKVGNLPAQRQLSLYQQTAS-----PI-----A 389  
QY 331 CNLFKLEYDGVFNKYTAQFDLNTKIGSRKIINRNYOFNYSLSNPYTNLNLTA-AVNSG 389  
Db 390 GTLNL--GTRPPYELGWK---RTGYDYMARNAADYSLAPEDVDLDFQAKLYVD 442  
QY 390 RQYKPGSKFTGWGLLKDFETYNNAKILDLNNTATFLPRE--TELQTLGFNY-----F 442  
Db 443 TQ-----DSDSTYSTSLID-NGYATRLRTYGAQAQNTSRFLAPGHDF 487  
QY 443 HNEYGKNRPEELGLFDPDODNGLYSYLGKFGDKGLLPQKSTIYQVAGSQYFNTFY- 501  
Db 488 RANYG-----LEFYDKATSSR-----QCMGCVTPAGNRV---ASLFANLTYD 530  
QY 502 -----FDAALKDILYRLNYSNTVGY-----RFGGEXTGYGSDDEK--RAFGENSPY 549  
Db 531 YDGLWLTBGLRLYDRYLRGTG--LSYPLDAKDGORYT---IDNPKALRLTCCSTTR 585  
QY 550 KK-HCNRSCTGYEVLKYGKRRANNHSVSADFGFMPFASYSRTHRMPNTOEMVFS 608  
Db 586 EDWDVDRDQGLSPTL-----AVAVRPGV-EMLELYTYGKSWRPPAITE----- 629  
QY 609 QIGDSGVHTA-----LKPERANTWQFGNTYKGLKODDGLGLKLVGRSRIDNYI 660  
Db 630 TLNGSAHSSSTQPNPLQPSRAMEVGVNQVQPDWLFEGDLRAKVGYDFKVDNYI 689  
QY 661 HNVYKGMWDLNGDIPSWYS-STGLAYTIQHRNFKDKVHKHGFELNLYDYGRRFTNLSYA 719  
Db 690 NLA-----IDRNKPLVQPSIGNAAVY--NLSKTRFRLGLEYNVADGVFADLYTT 740  
QY 720 YQ--KSTOPTN-----FSDASEPN-----NASKEDQKQGYGLSRVSA--LPRDY 761  
Db 741 HMIKNEFCNSKAWLGLRLYGDGSRGRNFYVEPDAAASDFVTCDDGTFQFGSAAYLPGR 800  
QY 762 GRLEVTGRLWGLKLTGAMRYFGKSIIRATAERYIDGTNGNT-SNFRQLGRKSIRKOTE 820  
Db 801 G-----SVTLGG--RAFORKLDAGVTVRFAPGYQDSSVPSNY-----P 836  
QY 821 TIARQP--LIFDYAAYPKKNLIFRAEVKNLFDORRYIDPDAGNDAATQRYSSFPDK 878  
Db 837 YLADWPXYTLDELASYKLTDSLTLRGSVENLTNRAVY-----874  
QY 879 KDEVTGNADKTLONGYKGTGSKSVLTFNFRGRFTLITMSYKF 921  
Db 875 -----VSYGETLANTL-----GRGRTVQGGVEYRF 899

RESULT 12  
Q9CKJ4  
ID Q9CKJ4 PRELIMINARY; PRT; 848 AA.  
AC Q9CKJ4;  
DT 01-JUN-2001 (TremBrel. 17, Created)  
DT 01-JUN-2001 (TremBrel. 17, Last sequence update)  
DT 01-DEC-2001 (TremBrel. 19, Last annotation update)  
DE HASR.  
GN HASR OR PM1622.  
OS Pasteurella multocida.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Pasteurella.

NCBI\_TaxID=747;  
[1]  
RN SEQUENCE FROM N.A.  
RC STRAIN=PM70;  
RA MEDLINE=21145866; PubMed=11248100;  
RM May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;  
RT "Complete genomic sequence of Pasteurella multocida pm70.";  
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465 (2001).  
DR EMBL; AE006199; AAK03706.1; -;  
DR InterPro; IPR000531; TonB\_boxC.  
DR Pfam; PF00593; TonB\_boxC; 1.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; UNKNOWN1.  
KW Complete proteome.  
SQ SEQUENCE 848 AA; 95909 MW; 5C6B28E913F1D583 CRC64;  
  
Query Match 7.3%; Score 357.5; DB 16; Length 848;  
Best Local Similarity 21.6%; Pred. No. 1.4e-11;  
Matches 207; Conservative 153; Mismatches 364; Indels 235; Gaps 49;  
  
QY 12 FYLMGVLYHYSAEDAGRAGEAQIQVLEDVHVAKRVPKDK-----KVFTDARAYS- 64  
Db 14 FLMLGLPTWAFSQAN-----LEKSTINKLETILVNESE-EKNKFDENLIKTYLSSGSYSY 67  
QY 65 -TRODI--FKSSENLNIVRSIPGAFTOODKSSGIVSLNIRGDSGFGRVNTWVDGITQTF 121  
Db 68 LSQSDISTFRGS-SVGDFLGPVGVIVGNKRNKSGALSNIIRGIANENRVPVVID---KGL 123  
QY 122 YSTSTDRAGAGSGQFCASVDSNFIAGLDVVKGSFSGSAGINSIAGSLANLRTLGVDVVO 181  
Db 124 QSVPSYQYAGSSYR--TYLDPDLISQVEIEKGPSLQMDATGATGGVYRVETLRWQDIIP 181  
QY 182 GNNTYGLLLKGLTGTNS-----TKGNAMAAIGARKWLESAGSVGLVYGHs----RRSV 230  
Db 182 QGKVMGRLKGLTMTNTVSPPPYTRGGYQT-----KYISK-----CLSNHTGLCQTQY 231  
QY 231 AONRVGSGGQHGIFG-----AEYLERKORYFV-ORG-----ALAFN 268  
Db 232 APNARYSHGFDLNANNYSLAFANKWQADLVAYAKRKQGNFVGRHGQPTVIESIEFE 291  
QY 269 SDSGKWERDLQROOKYKPYKNYNNQELQKYEIEGHDKSWHENLAPQYDITPDPs-SLKQ 327  
Db 292 EDS-----VEVKEPRVHEVEIGSLTFKENRSTLY--RGEALNLSQ 332  
QY 328 QSAGNLKLE-YDGY-----FNKYTAQFDLNTKI-----GSRKIINRNTQF 368  
Db 333 DNTSYLAKINYNVDHRLGLAYRHYHSRFGETMSILNFRAYGALQEGTEVKVD-SYHA 391  
QY 369 NYGLS-LNPYTNLNLTAAYNSGRQYKPGSKFT-----GHGLLKDFETY--NNAKILDL 419  
Db 392 NYSYNPTTPYVNLNNAVYFTDS-----DSSNFTPIEYGYSLSRHAHFLVSKOKGLSI 446  
QY 420 NNTATFRL-PRETELQTLGFNFHNEYGNK--PPEELGLFDPDQDNGLYS 470  
Db 447 ETSIFQLNDKPFSLKYGLAHSEYRIYOPNAQARVAKGYPEDA----IGP-----LYI 497  
QY 471 YLGRFKGDKGLL-----POKSTIVQAGSQYFNTFYDAAKKDIYRLNYSNTNTVGYRFG 526  
Db 498 RDGKRKWSAFVAANYPITSMLKADIGLRYLQSTIYD-----YIVRTERVNIIG 546  
QY 527 EYTGYYGSD-----DEFKRAFGENSPYKHKCNKSCGI-YEPVLKYYKKKRRANNHSVIS 580  
Db 547 AFVNPNGPGNIWVEKYKDVVHKQAPVKNKGMSPVIMLTFFPI-----NGVQI- 594  
QY 581 ADFGDYFMPFASYSRTHRMPNIQEMVFSQIGDSGVHTA-----LKPERANTWQFGENTYK 635  
Db 595 -----YTKYAEALRSPS---LFQATKGSWSATADNLEQLRPERAKWKEAGINLIFY 642  
QY 636 KGLLKQDDTLGLKLVGRSRIDNTYIHNVYKGMWDLNGDIPSWYSSTGLAYTIQHRNFKDK 695  
Db 643 ENLGGKNILGFLKAYFNRIKDYLTRSYSP-----KDKVTQT---INIQSAQPK-- 689  
QY 696 VHKHGFELNLYDGRFTNLSYAYQKSTQPTNFSDESNNASKEDQLKQGY-----GL 751

```
Db 690 -----GIELSAYDMGKFAKLAGTYTKT-----ECLTAE-----QAGKEQCSGYIYRNSL 739
QY 752 SRVSALPRDYGRLEVGIRWLNKLTLCGAMRYFKS-IRATAERYIDGTNGGNTSFRQ 810
Db 740 NN-AVPPRLNHLATGLRFLFKQKLDIGARYSYYSKRILVPLVLSAERFV-----NTSSIE- 791
QY 811 LGRSIRKOTETLARQPLIFDYAAEYKPKNLIFRAEYKVNLFDRYIDPLDAGNDAATOR 869
Db 792 -----WAYSLV-DLIYANYNSNKLKTWTMDVFNRYILDNWGLNTAPGR 838

RESULT 13
ID Q98L70 PRELIMINARY; PRT; 747 AA.
AC Q98L70:
DT 01-OCT-2001 (TREMBLrel. 18, Created)
DT 01-OCT-2001 (TREMBLrel. 18, Last sequence update)
DT 01-OCT-2001 (TREMBLrel. 18, Last annotation update)
DE HEME ACQUISITION PROTEIN, HASR.
OS MLR1155.
GN Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Iidesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mechizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL; AP002996; BAB48593.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 747 AA; 80102 MW; 4105DA001B73E752 CRC64;
```

```
Query Match 7.1%; Score 346; DB 16; Length 747;
Best Local Similarity 21.4%; Pred. No. 4.9e-11;
Matches 197; Conservative 120; Mismatches 303; Indels 302; Gaps 38;

QY 23 SYAEDAGRAGEAQIOVLEVDHVHAKRVPKDK-----KVFTDARAVSTRQDIFKS 72
Db 52 AHAQQAQTAPAGEQTDQ-----SKAQDEKAATAGATLLDKILVISRTGETAIESLAS 103
QY 73 SENLD-----NIVRSIPGAFTQDQKSSGIVSLNIRGDSGFGRVNTMVDGITO 119
Db 104 ASHVQDQQLAHRMATTPNEMLLGPVGATQADARRVSTSNIRGLQDFGRVAVIVGARQ 163
QY 120 TFFYSTSTADAGGSGGASVNSFNITAGLDVVVKGFGSGAGINSLAGSANRLTLGVDDV 179
Db 164 DFO-----RSDHGTOSTQYIDPELVKSVDIRGVPVANTYGSAGTGVVFFEDTKDAADF 216
QY 180 VQGNNTYGLLLKGLTGNSTNKGMAAIGARKWLESASVG-VLYGHSRVSVAQNRVGG 238
Db 217 LKPEETWASVTGYRESNG-KGWTTSGATGYREFNENWALGNIVYRN-----YDNK-DG 269
QY 239 GGQHGNFGAEYLERRKQRYFVQEGALKFNSDSKWERDLQROWKYKPKYNNQELQK 298
Db 270 GGVTVNGTGFVNL-----SGLKTS-----IRPTENSELKL 300
QY 299 YIEGHDKSWRENLAPOVDIPDPSSLKQOSAGNLFKLEYDGVFNKYTAQFR----- 350
Db 301 GWGSSDGNWDETSGM-----PVNDVLDKS-----NFTARYNTIDEDKSW 341
QY 351 -DLNTKIGSRKIIRNNYQFVGLSLNPNYTNLTAAVNSGR-----QKYPKGSKFTG 401
```

```
Db 342 LDLH-----INTSYN-----KTNLDLTSLVQKRFDPITGLPTVLPAQSOST- 383
QY 402 WLLKDPETNNAKILDNNNTATFRLPRETELQTTLGFNFHNEYKRNRPPEELGLPFDG 461
Db 384 -----FDVGTG--IDIWNTSRF-----ETGGIAHELTYGGDWVG 416
QY 462 PDQNGLYSVLGRFGDKGLLPQKSTIVQAGSQYENFVFD-----AALKKDIYRLNYS 516
Db 417 DVKTG-----GRAGGDSFTYPSGKRV--SGATVQDKLTWDMLEVIAGLRYDYSKDS 469
QY 517 T-NTVGYRFGGEYT-GYIGSDDEFKRAFGENSPYTKKHCRSCGIEYFVLKKGKRRANN 574
Db 470 TRETSGDRLSPRITGV-----SPF-----ESAGL--AGLQFYG----- 501
QY 575 HVSISADFGDYFMPFASYSTRHMPNIQEMYSQIGDSGV-----HTALKKPERANTWQ 628
Db 502 -----TYAEGYRSPSLTETLISGNHAGVTFFPFLPNLNPETGKTTE 544
QY 629 FGNTRYKGLLKODDTLGLKLVGRSRIKNIHNVYKGMWDLNGIDPSWSSSTGLAVTIO 688
Db 545 FGINYRQNDIFEPGDALRVAAAFHNVDYIDCVTILSPFAPGSGCP---FGPIPICFQ 601
QY 689 HRNFKDKVHKHGFELNLYDYGRFNTNLAYOKSTQPTNFDASESPNASKEDQLKQG 748
Db 602 YQNFQ-AQKIDGFELEGVDAGWGYAGLSAS-----ITNG 635
QY 749 YGLS-----RVSNALPRDYGRLEVGTEWLNKLTLCGAMRYFKSIRATAERYIDGTNG 802
Db 636 HTISYKGEADLATIPSSQVTAQLGRLEDKLTITVGVEQYNGKP-----K 681
QY 803 GNTSNFRQLGKRSIKOTETILARQPLIFDFVAAEYKPKNL--IFRAEYKVNLFDRYIDPLD 860
Db 682 GN-----AVAEDYTLVNAFASQATDNLKVDFRAD--NLFEDVKYANPLN 723
QY 861 AGND-----AATORY 870
Db 724 GSTTVAVYEPGITLKLAAATNRF 745

RESULT 14
Q93SH4 PRELIMINARY; PRT; 782 AA.
AC Q93SH4:
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE HMUR PROTEIN.
GN HMUR.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1105PC4;
RA Nienaber A., Hennecke H., Fischer H.M.;
RT "Discovery of a haem uptake system in the soil bacterium
RT Bradyrhizobium japonicum.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ311165; CAC38746.1; -.
SQ SEQUENCE 782 AA; 83713 MW; F8FC1EBFE568AA87 CRC64;
```

```
Query Match 6.4%; Score 311; DB 2; Length 782;
Best Local Similarity 20.9%; Pred. No. 4.3e-09;
Matches 192; Conservative 99; Mismatches 317; Indels 310; Gaps 39;

QY 30 RAGSEAOIQVLEVDHVHAKRVPKDKKVFVDARA----VSTRQDIFKSSNLDNIVRSIPG 85
Db 71 QAGSAAPVOTLDTITVAA---TKTRERADALAPVSSISLDQIQGLQPNRLSDFHVSVP 127
QY 86 AFTQDQKSSGIVSLNIRGDSGFGRVNTMVDGITTFTYSTDAGRAGGSSQFGASVDSNF 145
```

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:44:38 ; Search time 62.3 seconds  
(without alignments)  
1643.821 Million cell updates/sec

Title: US-09-762-926-4  
Perfect score: 4904  
Sequence: 1 MRSSFRLPFCFLMGVLMVLY.....SVLTNPARGRFLTMSYKF 922

Scoring table:

BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 747574 seqs, 111073796 residues

Total number of hits satisfying chosen parameters: 747574

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : A\_Geneseq\_032802.\*

- 1: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1980.DAT.\*
- 2: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1981.DAT.\*
- 3: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1982.DAT.\*
- 4: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1983.DAT.\*
- 5: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1984.DAT.\*
- 6: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1985.DAT.\*
- 7: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1986.DAT.\*
- 8: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1987.DAT.\*
- 9: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1988.DAT.\*
- 10: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1989.DAT.\*
- 11: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1990.DAT.\*
- 12: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1991.DAT.\*
- 13: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1992.DAT.\*
- 14: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1993.DAT.\*
- 15: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1994.DAT.\*
- 16: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1995.DAT.\*
- 17: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1996.DAT.\*
- 18: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1997.DAT.\*
- 19: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1998.DAT.\*
- 20: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA1999.DAT.\*
- 21: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2000.DAT.\*
- 22: /SIDS1/gcgdata/hold-geneseq/geneseq-emb1/AA2001.DAT.\*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	4904	100.0	922	21	A BASB024 outer me
2	4894	99.8	922	21	A BASB024 outer me
3	4779.5	97.5	921	21	A BASB024 outer me
4	4754	96.9	922	20	N. gonorrhoeae ant
5	4743	96.7	922	20	Neisseria gonorrh
6	4599.5	93.8	888	20	Neisseria meningit
7	4413	90.0	871	20	N. meningitidis st
8	2598	53.0	947	21	M. catarrhalis (AT
9	2514.5	51.3	913	21	Haemophilus antige
10	2489	50.8	918	21	Haemophilus antige
11	1976	40.3	393	20	Neisseria meningit

12	900	18.4	753	22	ABB52934	Escherichia coli p
13	900	18.4	753	22	ABB52968	Escherichia coli p
14	278	5.7	915	20	AAV507476	N.gonorrhoeae trans
15	278	5.7	915	21	AAV51770	N. gonorrhoeae str
16	278	5.7	915	21	AAV80374	N. gonorrhoeae str
17	269.5	5.5	908	21	AAV51768	N. meningitidis str
18	269.5	5.5	908	21	AAV80372	N. meningitidis str
19	262.5	5.4	971	21	AAV80372	Haemophilus somnus
20	261	5.3	909	20	AAV07477	N.meningitidis tra
21	260	5.3	791	17	AAV95566	N.meningitidis tra
22	259	5.3	790	17	AAV95565	N. meningitidis se
23	257	5.2	911	21	AAV51769	N. meningitidis str
24	257	5.2	911	21	AAV80373	N. meningitidis str
25	257	5.2	915	22	AAE12018	Neisseria meningit
26	255	5.2	908	15	AAV48220	N.meningitidis IM2
27	250	5.1	884	14	AAV34402	Sequence of low mo
28	245	5.0	912	16	AAV96969	Transferin recept
29	238	4.9	790	17	AAV95568	N. gonorrhoeae B h
30	237.5	4.8	792	17	AAV95567	N. meningitidis se
31	233	4.8	912	16	AAV77886	Bacterial transfer
32	233	4.8	912	18	AAV08959	Amino acid sequenc
33	233	4.8	912	18	AAV08961	Amino acid sequenc
34	233	4.8	912	19	AAV53044	H. influenzae str
35	233	4.8	912	19	AAV53046	H. influenzae str
36	233	4.8	912	21	AAV51689	H. influenzae type
37	233	4.8	912	21	AAV51691	H. influenzae type
38	233	4.8	912	21	AAV80355	H. influenzae type
39	233	4.8	912	21	AAV80357	H. influenzae type
40	232.5	4.7	913	16	AAV77884	Bacterial transfer
41	230.5	4.7	911	16	AAV77892	Bacterial transfer
42	230.5	4.7	911	18	AAV08966	Amino acid sequenc
43	230.5	4.7	911	19	AAV54124	H. influenzae str
44	230.5	4.7	911	21	AAV51780	H. influenzae non-
45	230.5	4.7	911	21	AAV80362	H. influenzae str

ALIGNMENTS

RESULT 1  
AAV69381  
ID AAV69381 standard; Protein; 922 AA.  
XX AC AAV69381;  
XX AC AAV69381;  
XX DT 19-JUN-2000 (first entry)  
XX DE A BASB024 outer membrane protein of N. meningitidis.  
XX KW BASB024; outer membrane protein; N. meningitidis infection;  
XX KW bacteremia; meningitis.  
XX OS Neisseria meningitidis.  
XX PN WO200011182-A1  
XX PD 02-MAR-2000  
XX PF 13-AUG-1999; 99WO-EP05989.  
XX PR 18-AUG-1998; 98GB-0018004.  
XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
XX Thonhard J; appl.  
XX WPI; 2000-224702/19  
XX N-PSDB; AA26158T

Novel polypeptides derived from the products of the BASB024 gene of Neisseria meningitidis, useful for inducing an immune response and producing antibodies useful for treating meningitis -

PS Claim 3; Page 88-91; 103pp; English.

XX The present sequence represents a BASB024 outer membrane protein of

CC Neisseria meningitidis. The BASB024 polynucleotide sequence was

CC first identified in the Incyte Pathoseq database containing

CC unfinished genomic DNA sequence of N. meningitidis. BASB024

CC polypeptides and polynucleotides are useful for generating an

CC immune response in an animal. Antibodies specific BASB024 polypeptides

CC are useful for treating N. meningitidis infection, which causes

CC bacteremia and meningitis.

XX

SQ Sequence 922 AA;

Query Match 100.0%; Score 4904; DB 21; Length 922;

Best Local Similarity 100.0%; Pred. No. 0;

Matches 922; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 MRSSFLKPKICFYLMGVMLYHHSHVAEDAGRAGSEAOQVLEDVHVHAKRVPKDKKVFDTDA 60

Db 1 mrssfrlkipicfyilmgvmlhyhshvaedagragseaqivledvhvakrvpdkkvvftda 60

Qy 61 RAVSTRQDIFKSSNLDNIVRSIPCAFTQODKSSGIYSLNIRGDSGFGFVNTWVDGIGTQT 120

Db 61 ravstrqdifkssnldnivrsipgafqtqdkssgislirgdsgrfvrntwvmdvgtqt 120

Qy 121 FYSTSDAGRAGSSQFASVDSNFIAGLDVVKGFSFGSAGINSLAGSANLRTLCGVDDVV 180

Db 121 fystsdagragssqfagsvdsnfiaagldvvkgsfsgsaginslagsanlrtlcvddvv 180

Qy 181 QGNNTYGLLLKGLTGNTSTKNMAAATGARKWLESAGSVGLYCHSRSSVAQNYRVGGGG 240

Db 181 qgnntygllllkgltgntstknmaaatgarkwlesagsvglvchsrssvaqnyrvgggg 240

Qy 241 QHIGNFAEYLERKQRFVVEGGLKNSNGKWERFORPYWKTWYQKYNDDPQELQY 300

Db 241 qhignfaeylerkqrfvveggllknsngkwerforpywktwyqkynddpqelqy 300

Qy 301 IEHDKSWRENLAPOYDITPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQFRDLNKTIGSR 360

Db 301 ieghdkswrenlapoyditiPIDPSLLKQOSAGNLFKLEYDGVFNKYTAQFRDLNKTIGSR 360

Qy 361 KIIRNQFNGLSINSYANLNLAAYNSGRQKPKSGFTGWLGLKDFETYNNAKLIDL 420

Db 361 kiirnfqfnysinsyanlnlaaynsgrqkpksgftgwlglkdfetynnaklidl 420

Qy 421 NNTATFRLPRETELQTTILGFNFYFNEYKKNRFPPEELGFFDGPQDNGLYSLGRFRGDK 480

Db 421 nntatfrlpretelqttilgfnyfneykknrfppeelgffdgppdnglyslgrfrgdk 480

Qy 481 GLLPKSTIVPAGSQFNTFYDAALKKDIYRLNYSNTVNVGYRFGGEYTGYSDDDEFK 540

Db 481 glpstkstivpagsqfntfydaalkkdiyrlnysntvnygyrfggeygyysdddefk 540

Qy 541 RAFGENSPYKHKCNQSGIYEPVLLKYGKKRANHVSISADFGDYFMPFASYSRTHRM 600

Db 541 rafgensptykhhcnqsgiyepvllkygkkranhvsisadfgdyfmpfasystrhrm 600

Qy 601 PNIQEMFYSQIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLVGYSRIDNY 660

Db 601 pniqemfysqigdsgvhtalkperantwqgfntykkgllkqddtlglklvgysridny 660

Qy 661 IHNVYKWDNLNGNIPSWSVSTGLAYTIQHRNFKDKVHKHGFLELNYDYGRFTNLSYA 720

Db 661 ihnyvkwndlngnipswsvstglaytiqhrnfkdvkhkgfellelndygrfftnlsya 720

Qy 721 YKSTQPTNFSDESPPNASKEDQLQGYCLSRVLSALPRDYGLEVGTFRWLGKHLILGG 780

Db 721 ykstpntnfsdesppnaskedqlqgyclsrvsalprdyglevgtfrwlgkhlilgg 780

Qy 781 AMRYFGKSIRATABERYIDGTNGNNTSNVROLGKRISIKQETTLARQPLIFDYAAYEPKK 840

Db 781 amryfgksirataberyidgtngnntsnvrolgkriskqettlarqplifdyaayepkk 840

Qy 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFPDKDEEYTCNADKTLCKNGYGGT 900

Db 841 nlifraevknlfdriryidpldagndaatqrysfpdkdkdeevtcnadtclngkyggt 900

Qy 901 SKSVLTNFAARGRTFLITMSYKF 922

Db 901 sksvltnfargrtflitmsykf 922

RESULT 2

AAV69380

ID AAY69380 standard; Protein; 922 AA.

XX

AC AAY69380;

XX

DT 19-JUN-2000 (first entry)

XX

DE A BASB024 outer membrane protein of N. meningitidis.

XX

KW BASB024; outer membrane protein; N. meningitidis infection;

XX

OS Neisseria meningitidis.

XX

PN WO200011182-A1.

PD 02-MAR-2000.

PF 13-AUG-1999; 99WO-EP05989.

PR 18-AUG-1998; 98GB-0018004.

XX (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.

XX

PI Thonnard J;

XX

DR WPI; 2000-224702/19.

DR N-PSDB; AAZ61580.

XX

PT Novel polypeptides derived from the products of the BASB024 gene of

PT Neisseria meningitidis, useful for inducing an immune response and

PT producing antibodies useful for treating meningitis -

XX

PS Claim 5; Page 83-86; 103pp; English.

XX

CC The present sequence represents a BASB024 outer membrane protein of

CC Neisseria meningitidis. The BASB024 polynucleotide sequence was

CC first identified in the Incyte Pathoseq database containing

CC unfinished genomic DNA sequence of N. meningitidis. BASB024

CC polypeptides and polynucleotides are useful for generating an

CC immune response in an animal. Antibodies specific BASB024 polypeptides

CC are useful for treating N. meningitidis infection, which causes

CC bacteremia and meningitis.

XX

SQ Sequence 922 AA;

Query Match 99.8%; Score 4894; DB 21; Length 922;

Best Local Similarity 99.8%; Pred. No. 0;

Matches 920; Conservative 1; Mismatches 1; Indels 0; Gaps 0;

Qy 1 MRSSFLKPKICFYLMGVMLYHHSHVAEDAGRAGSEAOQVLEDVHVHAKRVPKDKKVFDTDA 60

Db 1 mrssfrlkipicfyilmgvmlhyhshvaedagragseaqivledvhvakrvpdkkvvftda 60

Qy 61 RAVSTRQDIFKSSNLDNIVRSIPCAFTQODKSSGIYSLNIRGDSGFGFVNTWVDGIGTQT 120

Db 61 ravstrqdifkssnldnivrsipgafqtqdkssgislirgdsgrfvrntwvmdvgtqt 120

Qy 121 FYSTSDAGRAGSSQFASVDSNFIAGLDVVKGFSFGSAGINSLAGSANLRTLCGVDDVV 180

Db 121 fystsdagragssqfagsvdsnfiaagldvvkgsfsgsaginslagsanlrtlcvddvv 180

Qy 181 QGNNTYGLLLKGLTGSTKGNAMAAIGARKWLESASGVLYIGHSRSSVAQNYRVGGG 240  
 Db 181 qgnntyglllkgltgtnstkgnamaaigarkwlesasgvlyghsrrtwagnyrvsgg 240  
 Qy 241 QHIGNFGAEXLERKQRYFVQEGGLKNSNGKWERDQRPYWKYQKYNDRPQELQY 300  
 Db 241 qhignfgaeylerkryfvqegglknsngkwerdqrpywkyqkynndrpqelqy 300  
 Qy 301 IEGHDKSWRENLAPOYDITPIDPSLSKQOSAGNLFKLEYDGVFNKYTAQFRDLNKTIGSR 360  
 Db 301 ieghdkswrenlapqyditpidpslskqosagnlfkleydgvfnkytaqfrdlntkigr 360  
 Qy 361 KIINRNTQFNYSLSNSYANLNLAAYNSGRQKYPKSGKFTGWGLLKDFEYNNAKILDL 420  
 Db 361 kiinrnyqfnyslsnysanlnltaaynsgrqkypksgkftgwlldkfeynnakildl 420  
 Qy 421 NNTATFRLPRETELOTTILGFNFYFNEVYKGRNPFPEELGLFFEDGPDODNGLYSLGRFGDK 480  
 Db 421 nntatfrlpreteletlgtlgnfyfneygknrfpeelglffdgpdodnglyslgrfgdk 480  
 Qy 481 GLLPQKSTIVOPAGSQYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFFK 540  
 Db 481 gllpqkstivopagsqyfnfitydaalkkdiyrlnysntvgyrfgeygyysddefk 540  
 Qy 541 RAFGENSPYKHKHCNQCSCIYEPVLYKYGKRRANNHVSISADFGDYFMPFASYSRTHRM 600  
 Db 541 rafgensptykhhcnscgiyepvlykygkrrannhsvsisadfgdyfmpfasysrthrm 600  
 Qy 601 PNIQEMVFSQIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660  
 Db 601 pniqemvfsqigdsghvhtalkperantwqgfntyykkgllkqddtlglklvgyrsridny 660  
 Qy 661 IHNYGKWDNLNGNIPSWSVSTGLAYTIOHRNFKDKVHKHGFELNLYDYGRRFTNLSYA 720  
 Db 661 ihnygkwdnlngnipswsvstglaytiqhrnfkdvkhkhgfelnlydygrrftnlsya 720  
 Qy 721 YQKSTQPTNFSDESAPNASKEDQLQGGYLSRVSAIPRDYGRLEVTGRWLNKLLGG 780  
 Db 721 yqkstqptnfsdesapnaskedqlqgyysrvsaiprdyygrlevtrwlnklltg 780  
 Qy 781 AMRYFGKSIRATAERYIDGTNGGNTSNVROLGKRISIKQETTLARQPLIFDEYAAEPPK 840  
 Db 781 amryfgksirataeeryidgtnggntsnvrolgkriskqetltarqplifdeyaaepk 840  
 Qy 841 NLIFRAEYKLNIFDRYIDPLDAGNDAATQRYYSFDPKDEEVTCTNADKTLCKNGYGGT 900  
 Db 841 nlifraeyklnifdryidpldagndaatqryysfdpkdeevtctnadkctlcngkygt 900  
 Qy 901 SKSVLTNFARGTRFLTMSYKF 922  
 Db 901 sksvltnfargtrfltmsyaf 922

## RESULT 3

AA69382  
 ID AAY69382 standard; Protein; 921 AA.

AC AAY69382;

DT 19-JUN-2000 (first entry)

XX A BASB024 outer membrane protein of N. meningitidis.

XX BASB024; outer membrane protein; N. meningitidis infection;  
 KW bacteremia; meningitis.

XX Neisseria meningitidis.

XX W0200001182-A1.

XX 02-MAR-2000.

XX

PF 13-AUG-1999; 99WO-EP05989.  
 XX  
 PR 18-AUG-1998; 98GB-0018004.  
 XX  
 PA (SMIK ) SMITHKLINE BEECHAM BIOLOGICALS.  
 XX  
 PI Thonnard J;  
 XX  
 DR WPI; 2000-224702/19.  
 XX  
 DR N-PSDB; AAZ61582.  
 XX  
 PT Novel polypeptides derived from the products of the BASB024 gene of  
 PT Neisseria meningitidis, useful for inducing an immune response and  
 PT producing antibodies useful for treating meningitis -  
 XX  
 XX Claim 3; Page 91-95; 103pp; English.  
 XX  
 CC The present sequence represents a BASB024 outer membrane protein of  
 CC Neisseria meningitidis. The BASB024 polynucleotide sequence was  
 CC first identified in the Incyte Pathoseq database containing  
 CC unfinished genomic DNA sequence of N. meningitidis. BASB024  
 CC polypeptides and polynucleotides are useful for generating an  
 CC immune response in an animal. Antibodies specific BASB024 polypeptides  
 CC are useful for treating N. meningitidis infection, which causes  
 CC bacteremia and meningitis.  
 XX  
 SQ Sequence 921 AA;

Query Match 97.5%; Score 4779.5; DB 21; Length 921;  
 Best Local Similarity 97.9%; Pred. No. 0;  
 Matches 903; Conservative 7; Mismatches 11; Indels 1; Gaps 1;

Qy 1 MRSSFRLKPICFYLMGVMLYHHSYAEADAGRAGSEAOIQVLEDVHVAKRVPKDKVFTDA 60  
 Db 1 mrssfrikpicfyimgvtlyhysyaedagragseaqiqlvedvhwkkrvpkdkkvftda 60  
 Qy 61 RAVSTRQIDFKSSNNLDNIVRSIFCAFTQDKSGSIVSLNIRGDSGGFORVMTWVDGITQT 120  
 Db 61 ravstrqidfksennldnivrslpcaftqdksgsivslnirgdsfggrvmtvmdgitqt 120  
 Qy 121 FYSTSTAGRAGSGSOFASVDSNFIAGLDVVKGFSGSAGINSIAGSANLRTLGVDVV 180  
 Db 121 fyststagragsgsqfagsvdsnfiafldvvkgfsgsaginslagsanlrtlvgddvv 180  
 Qy 181 QGNNTYGLLLKGLTGSTKGNAMAAIGARKWLESASGVLYIGHSRSSVAQNYRVGGG 240  
 Db 181 qgnntyglllkgltgtnstkgnamaaigarkwlesasgvlyghsrrswagnyrvsgg 240  
 Qy 241 QHIGNFGAEXLERKQRYFVQEGGLKNSNGKWERDQRPYWKYQKYNDRPQELQY 300  
 Db 241 qhignfgaeylerkryfvqegglknsngkwerdqlgrqgkwykynndrpqelqy 299  
 Qy 301 IEGHDKSWRENLAPOYDITPIDPSLSKQOSAGNLFKLEYDGVFNKYTAQFRDLNKTIGSR 360  
 Db 301 ieghdkswrenlapqyditpidpslskqosagnlfkleydgvfnkytaqfrdlntkigr 359  
 Qy 361 KIINRNTQFNYSLSNSYANLNLAAYNSGRQKYPKSGKFTGWGLLKDFEYNNAKILDL 420  
 Db 361 kiinrnyqfnyslsnysanlnltaaynsgrqkypksgkftgwlldkfeynnakildl 419  
 Qy 421 NNTATFRLPRETELOTTILGFNFYFNEVYKGRNPFPEELGLFFEDGPDODNGLYSLGRFGDK 480  
 Db 421 nntatfrlpreteletlgtlgnfyfneygknrfpeelglffdgpdodnglyslgrfgdk 479  
 Qy 481 GLLPQKSTIVOPAGSQYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFFK 540  
 Db 481 gllpqkstivopagsqyfnfitydaalkkdiyrlnysntvgyrfgeygyysddefk 539  
 Qy 541 RAFGENSPYKHKHCNQCSCIYEPVLYKYGKRRANNHVSISADFGDYFMPFASYSRTHRM 600  
 Db 540 rafgensptykhhcnscgiyepvlykygkrrannhsvsisadfgdyfmpfasysrthrm 599

QY 601 PNQEMYSQIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660  
 Db 600 pniqemyfsqigdsghvhtalkperantwqgfntykkglkqddtlglklvgyrsridny 659  
 QY 661 IHNVYKQWDLNGNIPSWSGTGLAYTIOHRNFKDKVHKHGFELNVDYGRFTNLSYA 720  
 Db 660 ihnvygkwdlndgngipswsgtglaytiqhrnfkdvkhghfelnvdygrftnlsya 719  
 QY 721 YKSTQPTNFSASESPNNASKEDQLKQGYCLSRVSLPRDYGRLEVTGRLWGNKLTGG 780  
 Db 720 ykfstqptnfsdasespnnaskedqlkqgyclsrvsalprdygrlevtgrlwnkltl99 779  
 QY 781 AMRYFGKSIRATAERYIDGTNGTGNVSNRQLGKRSIKQTETLARQPLIFDFAAYEPKK 840  
 Db 780 amryfgksirataeeryidgtngtgnvnfrqlgkrsikqtetlarqplifdyaayepkk 839  
 QY 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFDPKDDEEVTNCNADKTLGNGKYGGT 900  
 Db 840 nlifraevknlfdriryidpldagndaatqryssfdbpkddedvtcnadktlcngkyggt 899  
 QY 901 SKSVLTNFARGRTFLITMSYKF 922  
 Db 900 sksvltnfargrtflitmsykf 921

RESULT 4  
 AAY38940  
 ID AAY38940 standard; Protein; 922 AA.  
 XX AC AAY38940;  
 XX DT 08-OCT-1999 (first entry)  
 XX DE N. gonorrhoeae antigen encoded by a variant ORF133.  
 XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 XX KW treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.  
 XX OS Neisseria gonorrhoeae.  
 XX PN WO9924578-A2.  
 XX PD 20-MAY-1999.  
 XX PF 09-OCT-1998; 98WO-IB01665.  
 XX PR 01-SEP-1998; 98GB-0019016.  
 XX PR 06-NOV-1997; 97GB-0023516.  
 XX PR 14-NOV-1997; 97GB-0024190.  
 XX PR 18-NOV-1997; 97GB-0024386.  
 XX PR 27-NOV-1997; 97GB-0025158.  
 XX PR 10-DEC-1997; 97GB-0026147.  
 XX PR 14-JAN-1998; 98GB-0000759.  
 XX PA (CHIR-) CHIRON SPA.  
 XX PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;  
 XX DR WPI; 1999-327407/27.  
 XX DR N-PSDB; AA212354.  
 XX PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
 XX PT diagnosis, treatment and prevention of infection  
 XX PS Claim 4; Page 480; 524pp; English.  
 XX CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis  
 XX CC and N. gonorrhoeae antigenic proteins. They are encoded by open  
 XX CC reading frames (ORFs) AA211972-212358. The antigenic proteins,  
 XX CC their fragments, their nucleic acids and antibodies are used for  
 XX CC diagnosis, prevention (as vaccines) or treatment of Neisseria  
 XX CC infections, such as meningitis, septicaemia and gonorrhea. Both  
 XX CC organisms are closely related. Fragments of the nucleic acids

CC are useful as hybridisation probes and antisense reagents.  
 XX SQ Sequence 922 AA;  
 Query Match 96.9%; Score 4754; DB 20; Length 922;  
 Best Local Similarity 96.3%; Pred. No. 0;  
 Matches 888; Conservative 16; Mismatches 18; Indels 0; Gaps 0;  
 QY 1 MRSSRLKPTCFYLMGVMLYHSHYAEDAGRAGSEAOIQVLEDVHKAKRYPKDKKVFDTDA 60  
 Db 1 mrsftrlkptcfylmgvmlhshyaedagragsaolqvledvkhakrpkdkkvvftda 60  
 QY 61 RAYSTRQDIFKSENENLDNIYRSIPGAFTQDKSSGIVSLMIRGDSGFRYNTWVDGITQT 120  
 Db 61 ravstrqdifkssenldniyrsipgaftqdkssgivslmirdsgfryntwvdgitqt 120  
 QY 121 FYSTSDAGRAGSSQPGASVDSNFTIAGLDVYKGSFSGSAGINSAGSALNRLTIGVDVV 180  
 Db 121 fystsdagragssgpgasvdsnftiagldvvykgsfsgsaginsagsalnrlt1gvdvv 180  
 QY 181 QGNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESAGSVGLYGHSSRSVAQNYRVGGGG 240  
 Db 181 qgnntygl1llkgl1gtgtnstkgnamaaigarkwlesgasvgl1ghssrgvaqnyrv9999 240  
 QY 241 QHIGNFGAYLERRKQRYFVOEGGLKFNSSGKWERDFQRPYWKTKYQKYNPDQELQY 300  
 Db 241 qhignfgeylerrkqgyfvgegglkfnagskwerdlqrqywktykwykyedpqelqy 300  
 QY 301 IEHDKSWRENLAPOYDITPIDPSSLKQOSAGNLKLEYDGVFNKYTAQFRDLNKTGSR 360  
 Db 301 iehdkswnrenlapoyditi1dps1llkqosagnl1kleydgvfnkytaqfrdlntr1gr 360  
 QY 361 KIINRNYQFNYSGLSNLANLNTAAYNSGRQKYPKSKFTGMGLLKDFETYNNAKILDL 420  
 Db 361 kiinrnyqfnysglslpntlnlntaaynsgrqkypkaktgwgllkdfetynnakildl 420  
 QY 421 NNTATFRLPRETELOTTLGFNFYHNEYGNRPFPEELGFFDGGDDNGLYSYLGFRFGDK 480  
 Db 421 nntatfrlpretelottl1gf1nf1hneygnrpf1ee1l1gff1d1g1dd1g1kl1v1gr1sr1dny 480  
 QY 481 GLLPQKSTIVQAGSOYFNTFYDAALKDIIYRLNYSNTVYRFGGEYTYGYGSDDEEK 540  
 Db 481 gl1pqkstivqagsy1fntfy1daa1lkddi1yrl1nys1ntv1yr1fgge1tygy1gs1d1eefk 540  
 QY 541 RAFGENSPYKKHCNOSCGIYEPVLKKGKRRANNHSSVSISADFGDYFMPFASYSRTHRM 600  
 Db 541 rafgenspaykehcdpscglyepvlkkygkkrannhsvsisadfgdyfmpfagysrthrm 600  
 QY 601 PNIOEMYFSQIGDSGVHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLVGYRSRIDNY 660  
 Db 601 pniqemyfsqigdsghvhtalkperantwqgfntykkglkqddtlglklvgyrsridny 660  
 QY 661 IHNVYKQWDLNGNIPSWSGTGLAYTIOHRNFKDKVHKHGFELNVDYGRFTNLSYA 720  
 Db 661 ihnvygkwdlndgngipswsgtglaytiqhrnfkdvkhghfelnvdygrftnlsya 720  
 QY 721 YKSTQPTNFSASESPNNASKEDQLKQGYCLSRVSLPRDYGRLEVTGRLWGNKLTGG 780  
 Db 721 ykfstqptnfsdasespnnaskedqlkqgyclsrvsalprdygrlevtgrlwnkltl99 780  
 QY 781 AMRYFGKSIRATAERYIDGTNGTGNVSNRQLGKRSIKQTETLARQPLIFDFAAYEPKK 840  
 Db 781 amryfgksirataeeryidgtngtgnvnfrqlgkrsikqtetlarqplifdyaayepkk 840  
 QY 841 NLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFDPKDDEEVTNCNADKTLGNGKYGGT 900  
 Db 841 nlifraevknlfdriryidpldagndaatqryssfdbpkddedvtcnadktlcngkyggt 900  
 QY 901 SKSVLTNFARGRTFLITMSYKF 922  
 Db 901 sksvltnfargrtflitmsykf 922



RESULT 5  
 AAY38939 standard; Protein; 922 AA.  
 XX AC AAY38939;  
 XX DT 08-OCT-1999 (first entry)  
 XX DE Neisseria gonorrhoeae antigen encoded by ORF133.  
 XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 XX treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.  
 XX OS Neisseria gonorrhoeae.  
 XX PN WO924578-A2-  
 XX PD 20-MAY-1999-  
 XX PF 09-OCT-1998; 98WO-IB01665.  
 XX PR 01-SEP-1998; 98GB-0019016.  
 XX PR 06-NOV-1997; 97GB-0023516.  
 XX PR 14-NOV-1997; 97GB-0024190.  
 XX PR 18-NOV-1997; 97GB-0024386.  
 XX PR 27-NOV-1997; 97GB-0025158.  
 XX PR 10-DEC-1997; 97GB-0026147.  
 XX PR 14-JAN-1998; 98GB-0000759.  
 XX PA (CHIR-) CHIRON SPA.  
 XX PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;  
 XX WPI; 1999-327407/27.  
 XX PT Proteins from Neisseria meningitidis and N. gonorrhoeae useful for  
 XX diagnosis, treatment and prevention of infection  
 XX PS Claim 4; Page 479; 524pp; English.  
 XX CC Amino acid sequences AAY38499-Y38944 represent Neisseria meningitidis  
 CC and N. gonorrhoeae antigenic proteins. They are encoded by open  
 CC reading frames (ORFs) AAZ11973-Z12358. The antigenic proteins,  
 CC their fragments, their nucleic acids and antibodies are used for  
 CC diagnosis, prevention (as vaccines) or treatment of Neisseria  
 CC infections, such as meningitis, septicaemia and gonorrhea. Both  
 CC organisms are closely related. Fragments of the nucleic acids  
 CC are useful as hybridisation probes and antisense reagents.  
 XX SQ Sequence 922 AA;  
 Query Match 96.7%; Score 4743; DB 20; Length 922;  
 Best Local Similarity 96.1%; Pred. No. 0;  
 Matches 886; Conservative 16; Mismatches 20; Indels 0; Gaps 0;  
 QY 1 MKRSRLKPCFYLGMVLMYHHSYAEDRAGRSEAQIQVLEDVHVAKRVPKDKVFDTDA 60  
 Db 1 mrsstrlkpicfylmgvmlhhsyaedragragseaqiqlvedvhvakrvpdkkvtfda 60  
 QY 61 RAVSTRQDIFKSSNLDNIVRSIPGAFQQDKSSGIVSLNIRGDSGFGRVNTWVDGITQT 120  
 Db 61 ravstrqdvfksgenldnivrslpgaftqqdkssgivslnirgdsfgfrvntwvdgitqt 120  
 QY 121 FYSTSTDRAGRGSSQFGASVDSNFIAGLDVVKGFSGSAGINSIAGSANLRLGLVDDVV 180  
 Db 121 fyststdegragrgssqfgasvdsnfiagldvvkgfsfsaginslagsanrlrlgvddvv 180  
 QY 181 QGNNTYGLLLKGLGTNSTKGNMAAIGARKWLESASVGLVYGHSSRVAQNVRVGGGG 240  
 Db 181 qgnntygllyllkglgtnstkgnmaaiagarkwlesasvglvghssrvgvaqnvrv999g 240

QY 241 QHIGNFGAEYLERRKQRYFVQEGGLKFNNSCKWERDORPYTKWKYKYNDOELQKY 300  
 Db 241 qhignfgeeylerrrkqgyfvqeggllkfnagskwerdlrgywtkwykkyedpqlqky 300  
 QY 301 IEHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEVDGVFNKYTAQFRLDNLTKIGSR 360  
 Db 301 ieehdkswrenlapoyditpidpsglkqqsagnllnleydgvfnkyaqfdrldnlntrigr 360  
 QY 361 KIINRYQFNGLSLSNYANLNTAAYNSGROKYPKGSKFTCGLLKDFETYNNAKILDL 420  
 Db 361 kiinryqfngylslnpytnlntaaynsgrqkypkgakftgwgllkdfetynnakildl 420  
 QY 421 NNTATFRPRETELOTLGCFNYPHNEYGKNRPEELGIFFDGPDODDNGLYSVLGRFKGDK 480  
 Db 421 nntatfrpretelgtlgfnyfhneygnrpfpeelgiffdgpddonglysvlgrfkdk 480  
 QY 481 GLLPOKSTIVQAGSQYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTYGSDDEFK 540  
 Db 481 glpdkstivqagsyfnfyfdaalkkdiylrlnystnainyrfggytygysenefk 540  
 QY 541 RAFGENSPYKKHCNOSCGIYEPVLKKYKGGKRRANNHSVISADFGDYFMPFASYSSTRHM 600  
 Db 541 rafgenspaykehcdpscglyepvlkkygkkrannhsvisadfgdyfmpfagysrthrm 600  
 QY 601 PNIOEWFQSOIGDSGVHTALKPERANTWQGFNTVKKGLKODDTLGLKLVGRSRIDNY 660  
 Db 601 pniqemyfsgiqdsgvhtalkperantwqgfntvkkglkqddalglkivgyrridny 660  
 QY 661 IHNVYKWMWDLNCGNIPSWVSSTGLAYTTQHRNFKDKVHKHGFLELNDYDGRFFTNLSYA 720  
 Db 661 ihnvvgkwwdlndgipswvgstglaytirhrnfkdvkhkhgfelelnydygrfftnlsya 720  
 QY 721 YOKSTQPTNFDASEPNNAKEDQLKQYGLSRVSALPRDYGRLEVGTRWLGKMLTLOG 780  
 Db 721 ygstqptnfdasespnnaekedqlkqyglsrvsalprdygrlevgrtwlgmklitlog 780  
 QY 781 AMRYFGKSTRATAERYIDGTNGGNTSNVRQLGKRSIKOTETLAPQLIFDFVAAVEPKK 840  
 Db 781 amryfgksirataeryidgtngntsnvrqlgksikotetlarqplifdfvayayepkk 840  
 QY 841 NLIFRAEYKLNFDRIYIDPLDAGNDAATQRYYSFDPKDKDEEVTGNADKTLNCKYGGT 900  
 Db 841 nlifraevknlfdrdyidpldagndaatqryyyssfdpkdkdedvtcnadktlcnkyggt 900  
 QY 901 SKSVLTNTARGTFLITMSYKF 922  
 Db 901 sksvltntargtrflmtmsykf 922  
 RESULT 6  
 AAY38937  
 ID AAY38937 standard; Protein; 888 AA.  
 XX AC AAY38937;  
 XX DT 08-OCT-1999 (first entry)  
 XX DE Neisseria meningitidis antigen encoded by a partial ORF133.  
 XX KW Neisseria meningitidis; Neisseria gonorrhoeae; antigen; vaccine;  
 XX treatment; Neisseria infection; meningitis; septicaemia; gonorrhea.  
 XX OS Neisseria meningitidis.  
 XX PN WO924578-A2-  
 XX PD 20-MAY-1999-  
 XX PF 09-OCT-1998; 98WO-IB01665.  
 XX PR 01-SEP-1998; 98GB-0019016.  
 XX PR 06-NOV-1997; 97GB-0023516.  
 XX PR 14-NOV-1997; 97GB-0024190.

PR 18-NOV-1997; 97GB-0024386.  
 PR 27-NOV-1997; 97GB-0025158.  
 PR 10-DEC-1997; 97GB-0026147.  
 PR 14-JAN-1998; 98GB-0000759.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;  
 XX  
 DR WPI: 1999-327407/27.  
 DR N-PSDB; AAZ12352.  
 XX  
 PT Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for  
 PT diagnosis, treatment and prevention of infection  
 XX  
 PS Claim 4; Page 474; 524pp; English.  
 XX  
 CC Amino acid sequences AAY38499-Y38944 represent *Neisseria meningitidis*  
 CC and *N. gonorrhoeae* antigenic proteins. They are encoded by open  
 CC reading frames (ORFs) AAZ11972-212358. The antigenic proteins,  
 CC their fragments, their nucleic acids and antibodies are used for  
 CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*  
 CC infections, such as meningitis, septicaemia and gonorrhea. Both  
 CC organisms are closely related. Fragments of the nucleic acids  
 CC are useful as hybridisation probes and antisense reagents.  
 XX  
 SQ Sequence 888 AA;

Query Match 93.8% Score 4599.5; DB 20; Length 888;  
 Best Local Similarity 97.8% Pred. No. 0;  
 Matches 869; Conservative 7; Mismatches 12; Indels 1; Gaps 1;

QY 34 EAQIOVLDDVHVAKRVPKDKKVFVDARAVSTRQDIFKSSSENLNIVRSIPGAFQODKS 93  
 DB 1 eadqivldvnhvakrvpkdkkvtadaravstrqdfksssenlndivrsipgafqodks 60  
 QY 94 SGIVSLNIRGDSGFRVNTMVDGITFTYSTDAGRAGSSQFGASVDSNFIAGLDVVK 153  
 DB 61 sgivslnirgdsgrvntmvdgitftystdagragsqfgasvdsnfiagldvdk 120  
 QY 154 GSFGSGAGINSLAGSANRLTGLVDVVOGNTTYGLLLKLTGTNSTKGNMAAIGARKWL 213  
 DB 121 gsfgsaginslagsanrltglvddvvgnttyglllkgltgntskgnmaaaigarkwl 180  
 QY 214 ESGASGVLYGHRRSVAQNYRVGGGQHGIGNFGAEYLERRKQRYFVQEGGLKFNSNGK 273  
 DB 181 esgasgvlyghrrsvaqnyrvgggqhignfgaeylerrkaryfvqegalkfnsdsk 240  
 QY 274 WERDFORPYWKTWYOKYNDPQELQYIEGHDKSWRENLAPOYDITPIDPSSLKQOSAGN 333  
 DB 241 werdlgrqwkypkyknynn-qelqyieehdkswrenlxpyditypidpsslkqqsagn 299  
 QY 334 LFKLEVDGVFNKYTAOFRLNTKIGSRKTIINRYQFNGLSLNYSANLNTAAYNSGRK 393  
 DB 300 lfkleydgvfnkyltagfrldntkigsrkliinryqfnnglslnpylnlntaaynsgrk 359  
 QY 394 YPKGSFTGWLGLKDFETYNNAKILDLNNTATPRLPRETELQTLTGTFNHYGKGRNP 453  
 DB 360 ypgskftgwlglkdfetynnakildnntatprlpretelqtltgtnfhygkgnrtp 419  
 QY 454 EEGLGFDPDQNGLYSYLGRPFKQGLLPKSTIVPAGSQYFNTFFDAALKKDIYR 513  
 DB 420 eeiglfdpdpdnglysyylgrpfkqglilpkstivpagsqyfnfdaalkkdiyr 479  
 QY 514 LNYSTNTVGVFGEYTGYYGSDPEKRAFGENSPYKHCNQSCGIIYEPVLKKYKKRA 573  
 DB 480 lnystntvgvfgeytgyygdsdekrafgenspykchcnscgiiyepvlkkykkra 539  
 QY 574 NNHSVSTISDFGYFMPFASYSYTHRMPIQEMFYFSQIGDSGVHTALKPERANTWQFGN 633  
 DB 540 nnhsvisadfgdyfmpfasysythrmpiqemfyfsqigdsghvhtalkperantwqfgn 599

QY 634 TYKKGLLKQDDTLGLKLVGYSRIDNYIHNVYKWWDLNIPSWSSVSTGLAYTIQHRNF 693  
 DB 600 tykkgllkqddtlgllklyvysridnyihnvgykwwdlngdipswssvstglaytiqhrnf 659  
 QY 694 KDKVHKHGFELNLYDYGRRFTNLNLSYAYOKSTOPTNFSASESPNNASKEQDLKQGYGLS 753  
 DB 660 kdkvkhghfelelnlydygrfftnlsyayqkstoptnfsasespnnaskedqikqgygls 719  
 QY 754 RVSALPRDYGRLEVGTRWLGKLTGLGAMRYFKSIRATAERYIDGTNGGNTSNVRQLG 813  
 DB 720 rvsalprdygrlevgrtwlgkltlglgamryfgksirataeryidgtnggntsnvrqlg 779  
 QY 814 KRSIKQTEILARQPLIFDFAAYEPKKNLIFRAEVKNLFDERRVIDPLDAGNDAATQRYYS 873  
 DB 780 krsikqtetlarqplifdfaayepkknlifraevknlfdrvidpldagndaatqryys 839  
 QY 874 SEDPKDKDEVTNCADKTLGNGKYGTSKSVLTNFAARGRTFLITMSYKF 922  
 DB 840 sdfpkdkdvtcnadkltcngkygtsksvltnfargrtflitmsykf 888  
 RESULT 7  
 AAY38938  
 ID AAY38938 standard; Protein; 871 AA.  
 XX  
 AC AAY38938;  
 XX  
 DT 08-OCT-1999 (first entry)  
 XX  
 DE N. meningitidis strain A antigen encoded by a partial ORF133.  
 KW *Neisseria meningitidis*; *Neisseria gonorrhoeae*; antigen; vaccine;  
 KW treatment; *Neisseria* infection; meningitis; septicaemia; gonorrhea.  
 XX  
 OS *Neisseria meningitidis*.  
 XX  
 PN WO9244578-A2.  
 XX  
 PD 20-MAY-1999.  
 XX  
 PF 09-OCT-1998; 98WO-IB01665.  
 XX  
 PR 01-SEP-1998; 98GB-0019016.  
 PR 06-NOV-1997; 97GB-0023516.  
 PR 14-NOV-1997; 97GB-0024190.  
 PR 18-NOV-1997; 97GB-0024386.  
 PR 27-NOV-1997; 97GB-0025158.  
 PR 10-DEC-1997; 97GB-0026147.  
 PR 14-JAN-1998; 98GB-0000759.  
 XX  
 PA (CHIR-) CHIRON SPA.  
 XX  
 PI Grandi G, Masignani V, Pizza M, Rappuoli R, Scarlato V;  
 XX  
 DR WPI: 1999-327407/27.  
 DR N-PSDB; AAZ12353.  
 XX  
 PT Proteins from *Neisseria meningitidis* and *N. gonorrhoeae* useful for  
 PT diagnosis, treatment and prevention of infection  
 XX  
 PS Claim 4; Page 477; 524pp; English.  
 XX  
 CC Amino acid sequences AAY38499-Y38944 represent *Neisseria meningitidis*  
 CC and *N. gonorrhoeae* antigenic proteins. They are encoded by open  
 CC reading frames (ORFs) AAZ11972-212358. The antigenic proteins,  
 CC their fragments, their nucleic acids and antibodies are used for  
 CC diagnosis, prevention (as vaccines) or treatment of *Neisseria*  
 CC infections, such as meningitis, septicaemia and gonorrhea. Both  
 CC organisms are closely related. Fragments of the nucleic acids  
 CC are useful as hybridisation probes and antisense reagents.  
 XX  
 SQ Sequence 871 AA;

Query Match 90.0%; Score 4413; DB 20; Length 871;  
Best Local Similarity 95.6%; Pred. No. 0;  
Matches 833; Conservative 3; Mismatches 35; Indels 0; Gaps 0;

QY 52 KDKKVFTRAVSTRODIFKSSNLDNIVRSIPGAFQODKSSGIVSLNTRGDSGFRVN 111  
DB 1 kdkkvtaravstrdifiksenldnivrxipgaftxqkssgsvslntrxdsgfgrvn 60

QY 112 TMVDGITQTFYSTDAGRAGSSQFCASVDSNFIAGLDVVKGSAGINSLAGSANL 171  
DB 61 tmvdgitxtfststdagragssqfgasvdsnfisagldvkvksgsaginslagsanl 120

QY 172 RTLGVDVVGNNYGLLGLTGTNSTKGNAMAAIGARKWLESASVGLYCHSRSA 231  
DB 121 rtlxvdvvgvngxytllgltgtntsgnmaaiarkwlesasvglvghsrsva 180

QY 232 QNYRVGGGGHGNFGAEYLERRKQRYFVQEGGLKFNNSGKWERDFQRPYKTKWQY 291  
DB 181 qnyrvvggghgnfgaeylerrkqryfegegglkfnnsqkwerdfqsywtkwyqy 240

QY 292 NDPQELQKYTEGHDKSWRENLAPOYDITPIDPSLQKQASAGNLFKLEYDGVFNKYTAQFR 351  
DB 241 dapqelqkyteghdksrenlapoyditypidpsllkqkqasagnlfkleydgvfnkytaqfr 300

QY 352 DLNWKIGSRKLIINRNQFNYGLSLNSYANLNLTAAYNSGKQKYPKSKFTGWLLKDFET 411  
DB 301 dlntkigrkiliinrnqfnyglslnsylnpytnlntaaynsgrkypkskftgwlkdfet 360

QY 412 YNNAKILDLNNTATFRLPRTELOTLTGNYFHNEYKGNRPFPEELGLFFDGPQDNGLYS 471  
DB 361 ynnakildlntatftrlprtelotltgtnyfhneykgnrpfpeelglffdgpxdnglys 420

QY 472 YLGRFGDKGLLPQKSTIVOPAGSQYFNTFYDAALKDIYRLNYSNTVYRFGSEYTG 531  
DB 421 ylgrrfgdkglpqlkstiropagsyfnfntfydaalkdiylrlnysntvyrfgseytg 480

QY 532 YGSDDEFKRAFGNSPTYKHKCNQSGIYEPVLKYGKRRANNHVSISADGDFPMPF 591  
DB 481 ygdsdefkrafngsptykxhcnsgciyepvlkyygkrrannhvsisadgdfpmpf 540

QY 592 ASYSRTHRMENIGEMFYFSQIGDSGVHTALKPERANTWQGFNTYKKGLLKODDTLGLKLV 651  
DB 541 asysrthrmnigemyfsqigdsghvhtalkperantwqgfntykkglkoddttlglklv 600

QY 652 GYRSRIDNYIHNHYGKWDNLGNIPSWSTGLAYTIOHRNFKDKVHKHGFELNLYDYG 711  
DB 601 gysridnyihnygkwdnlgnipswstglaytiqhrnfkdvkhkhgfelenlydyx 660

QY 712 RFTNLSYAYOKSTQPTNFSDESNNASKEDQLKOGYGLSRVSALPRDYGRLEVGTWR 771  
DB 661 rftnlsyayokstqptnfsdesnnaskedqlkgyglsrvsalprdygrlevgtwr 720

QY 772 LSNKLTGGAMRYFGKSIRATAERYIDGTNGTNSVRQLGKRSIKQTEFLARQPLIFD 831  
DB 721 lgnkltggamryfgksirataeryidxtngxtsnfrqlgkrsixqteflarqplifd 780

QY 832 FYAAYEPKMLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDXDEEVTCTNADKT 891  
DB 781 fyaayepkmlifraevknlfdriydpdldagnndaatqryysfdpkdxdeevtctndkt 840

QY 892 LCNKGYGTSKSLTNFARGRTFLITMSYKF 922  
DB 841 lcnkgygtsksltnfargrtflitmsyxf 871

RESULT 8  
AAY44428  
ID AAY44428 standard; Protein; 947 AA.  
XX  
AC AAY44428;  
XX  
DT 22-MAR-2000 (first entry)

XX M. catarrhalis (ATCC 43617) BASB021 polypeptide.  
DE BASB021; HasR; outer membrane haem-binding protein; sinusitis;  
XX otitis media; pneumonia; nosocomial infection; auditive nerve damage;  
KW delayed speech learning.  
KW Moraxella catarrhalis.  
OS  
XX WO9964602-A2.  
XX PN  
XX PD 16-DEC-1999.  
XX PF 31-MAY-1999; 99WO-EP03824.  
XX PR 09-JUN-1998; 98GB-0012440.  
XX PA (SMTK) SMITHKLINE BEECHAM BIOLOGICALS.  
XX PI Thonnard J;  
XX MPI: 2000-116545/10.  
DR N-PSDB; AA229682.  
XX New isolated Moraxella catarrhalis BASB021 polynucleotides, used to  
PT develop products for the diagnosis, prevention and treatment of  
PT infections causing e.g. otitis media -  
XX Claim 1; Page 81-84; 87pp; English.  
XX The present sequence is BASB021 polypeptide, which shows sequence  
CC homology with Serratia marcescens HasR outer membrane haem-binding  
CC protein. It is encoded by DNA from Moraxella catarrhalis strain Mc2931  
CC (ATCC 43617). BASB021 polynucleotides and polypeptides may be used for  
CC prognosis, staging of disease, determining response to drug treatment  
CC and drug screening. They can be used in vaccines for generating an immune  
CC response and for preventing or treating Moraxella infections which may  
CC cause otitis media, pneumonia, sinusitis, nosocomial infections and  
CC invasive diseases, auditive nerve damage, delayed speech learning,  
CC infection of upper respiratory tract and inflammation of the middle ear.  
CC Anti-BASB021 antibodies can be used to diagnose and treat Moraxella  
CC infections.  
XX Sequence 947 AA;  
SQ

Query Match 53.0%; Score 2598; DB 21; Length 947;  
Best Local Similarity 55.4%; Pred. No. 5.3e-179;  
Matches 501; Conservative 145; Mismatches 234; Indels 24; Gaps 10;

QY 39 VLEDVHVAKAR-VPKDKKVFTRAVSTRODIFKSSNLDNIVRSIPGAFQODKSSGIV 97  
DB 48 ildevvvtatngtkkqkpfkasatsrenvfnasenaialvrsypgafqdkksglv 107

QY 98 SLNIRGDSGFRVNTWVDGITQTFYSTDAGRAGSSQFCASVDSNFIAGLDVVKGSFS 157  
DB 108 slnvrsgdsgfransmvgtvtqtfyststdagrggtsqfgavidgnfiagvelnkgsfn 167

QY 158 GSAGINSLAGSANLRTLGVDVVGNNYGLLGLTGTNSTKGNAMAAIGARKWLESAG 217  
DB 168 gkgvntifgsanfrtlnadddvdkdknfgiakgtgkntdknfmلاغgrgwlndg- 226

QY 218 SVGLYCHSRSAQNYRVGGGGHGNFGAEYLERRKQRYFVQEGGLKFNNSGKWERD 277  
DB 227 sisalyayshkdisqykvggggthignvgdillskqkvafakehaltynearsrwdk 286

QY 278 ---FQRPYKTKWQY-----NDPQELQKYTEGHDKSWRENLAPOYDITPID 322  
DB 287 ltkldketgkplwdrkyfggkcygigldtkekfdeyvadvkqgqwkbgakeysitpid 346

QY 323 PSSLKQOSAGNLFKLEYDGVFNKYTAQFRDLNWKIGSRKLIINRNQFNYGLSLNSYANL 382  
DB 347 italnqtskshlakirynndtsdvglqrkmdttlgsrrisndnyqidaaynpneidlk 406







```
SQ Sequence 753 AA;

Query Match 18.4%; Score 900; DB 22; Length 753;
Best Local Similarity 28.7%; Pred. No. 1.9e-56;
Matches 272; Conservative 130; Mismatches 299; Indels 248; Gaps 32;

Qy 16 GVML-----YHSHYAE--DAGRAGSEAIQVLEDDVHVHAKRVPKDKVFTDARAVSTRQD 68
Dy 11 gilllscgaysisektnsdkga-aefspl-----vsvgkttsegealektgatssr-t 65
Qy 69 IFKSSNDLIVRSIPGAFTQODKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYSTSTDA 128
Dy 66 tdklqslatvrmptgttdipgqgaivsnirgmsfgfrvntmvdgtdtsgfytstsg 125
Qy 129 GRAGGS--SQFGASVDSNFIAGLDVVKGFSFGSAGINSLAGSANLRTLGVDVVGNNY 186
Dy 126 ttthgstnmagvliidpnllvavdvrgdssgseginalagsamrtigvddvifngnty 185
Qy 187 GLLKGLTGTNSTKGNMAAIGARK--WLESGASGVLYGHRSRVAQNYRVGGGQHIG 244
Dy 186 glrsfvgnglrgsmialgkgsdaftdtg-sigvmaavsgssvsnfngsg----- 239
Qy 245 NFGAEYLERRKQRFVQBGGLKFNSSGKWRDQRPVWKTWKQYKNDPOELOKYIEGH 304
Dy 240 -----inskefydk-----ymknqpsqlyk----- 261
Qy 305 DKSRENLAPOYDITPIDPSSLKOOSAGNLPKLEYGVFNKYTAQFRDLNLTIGSRKIIN 364
Dy 262 -----mdirpde-----fnfselsartyenknfrdrts 290
Qy 365 RNYQFNGLS--LNSYANLNTAAVNSGROKYPKGSKFTGWLKDDFETYNNAKILDLNNT 423
Dy 291 ddyikyhytfselidnvtastsrngkyrdsltyf-----ktsaqnrdsaldnnt 346
Qy 424 ATFRLPRETELQTTLGFNYFHNKYKRNFPPELGLFDDGPDQDNGLYSLGRFKDKGLL 483
Dy 347 srftv-aadndiefmlgsklmrty-----drtihsaagdphkandesi 387
Qy 484 POKTIVQAGSQVNTFYFDAAALKDI-----YRLNVTSTNVGVRFGEYTYGYSBDEF 539
Dy 388 --ennfpagsgqdisalytglktrgiweadfnlntnri-----tg----- 429
Qy 540 KRAFGNSPTVKKHCNOS--C--GIVEPVLKYYKKRANHSVSIADFGDYPFNPASY 594
Dy 430 -----ykpacdsrvicvpggsydidckeg-----fnpsvqlsaqvtwlpqfigy 475
Qy 595 SRTHRMNIQMYFSQIGDSGVHTALPERANTWQFGFNTYKGLLQDDDTLGLKLGYR 654
Dy 476 sksmrapniqemffnsggaamnpflkperatwqagfnidrdllveqdalrfkalayr 535
Qy 655 SRIDNYIHN-----VYKKW-----DLNGNIPSWSVSTGLAYTIQHRNF 693
Dy 536 sriqnyisylvcsgrkcselpvigngwegisdeysdmnyiynsa----- 584
Qy 694 KDKVHKHFELELNDYGRFPTNLSYAKSTQPTNDSASESPNNASKEDQLQGYCLS 753
Dy 585 sdvak-gfelemdydgafagrfsfqgqcdqptsasth-----fgag 628
Qy 754 RVSALPRDYGLEVGRWLGNKLTGGAMRYFGKSIRATAEERYIDGTNGGNTSNVRQLG 813
Dy 629 ditelprkymtldtgrvffdnalltltiikytkarrlspdfeqdehtga----- 678
Qy 814 KRSIKQETTLARQLIFDFYAAEPKKNLIPRAEVKNLFDPRYIDPDLADNDAATQRYYS 873
Dy 679 --iikq--dlpqiptlidltycyeynrltllksvqnlmnrdisenakln----- 725
Qy 874 SFDPKDDEEVTCHADKTLCKNGKGGTSGKSVLTNFAGRTEPLITMSYKF 922
Dy 726 -mmpglgdehpnans-----argrtwifggdirtf 753

RESULT 13
```

```
ABB52968
ID ABB52968 standard; Protein; 753 AA.
XX
XX ABB52968;
XX
XX 11-FEB-2002 (first entry)
DT
XX
XX Escherichia coli polypeptide SEQ ID NO 1277.
DE
XX
XX Escherichia coli; B2/D+A-; antiinflammatory; antibacterial;
KW immunosuppressive; extra-intestinal infection; phylogeny; meningitis;
KW systemic infection; non-diarrhoeal infection; septicemia;
KW pyelonephritis; antibiotic resistance.
XX
XX Escherichia coli.
OS
XX
XX WO200166572-A2.
PN
XX
XX 13-SEP-2001.
PD
XX
XX 12-MAR-2001; 2001WO-EP03445.
PF
XX
XX 10-MAR-2000; 2000FR-0003145.
PR
XX
XX 02-FEB-2001; 2001FR-0001449.
PR
XX
XX (INRM ) INSERM INST NAT SANTE & RECH MEDICALE.
PA
XX
XX Bingen E, Bonacorsi S, Clermont O, Nassif X, Tinsley C;
PI
XX
XX WPI; 2001-550253/61.
DR
XX
XX A library of DNA fragments of Escherichia coli strains for the
PT phylogenetic determination of a given strain comprises polynucleotides of
PT nature B2/D+ A- -
PT
XX
XX Example 6; Fig 6; 646pp; English.
PS
XX
XX The invention relates to a library of DNA fragments of Escherichia coli
CC strains comprising polynucleotides (ABA88577-ABA88729 and ABA89533)
CC and encoded proteins (ABB52459-ABB52919 and ABB52954-ABB53094) of nature
CC B2/D+A-. The polynucleotides have potential antiinflammatory,
CC antibacterial and immunosuppressive activity as part of pharmaceutical
CC compositions used to treat, palliate or prevent extra-intestinal E. coli
CC infections. The polypeptides are useful for determining the phylogenic
CC group of a given E. coli strain. These polypeptides can detect and treat
CC an undesired development of E. coli, particularly an extra-intestinal
CC infection that include systemic and non-diarrhoeal infections such as
CC septicemia, pyelonephritis and meningitis this is particularly
CC advantageous as bacterial resistance is increasing with the more
CC frequent use of broad spectrum antibiotics.
XX
XX Sequence 753 AA;
SQ
```

```
Query Match 18.4%; Score 900; DB 22; Length 753;
Best Local Similarity 28.7%; Pred. No. 1.9e-56;
Matches 272; Conservative 130; Mismatches 299; Indels 248; Gaps 32;
```

```
Qy 16 GVML-----YHSHYAE--DAGRAGSEAIQVLEDDVHVHAKRVPKDKVFTDARAVSTRQD 68
Dy 11 gilllscgaysisektnsdkga-aefspl-----vsvgkttsegealektgatssr-t 65
Qy 69 IFKSSNDLIVRSIPGAFTQODKSSGIVSLNIRGDSGFGRVNTMVDGITQTFYSTSTDA 128
Dy 66 tdklqslatvrmptgttdipgqgaivsnirgmsfgfrvntmvdgtdtsgfytstsg 125
Qy 129 GRAGGS--SQFGASVDSNFIAGLDVVKGFSFGSAGINSLAGSANLRTLGVDVVGNNY 186
Dy 126 ttthgstnmagvliidpnllvavdvrgdssgseginalagsamrtigvddvifngnty 185
Qy 187 GLLKGLTGTNSTKGNMAAIGARK--WLESGASGVLYGHRSRVAQNYRVGGGQHIG 244
Dy 186 glrsfvgnglrgsmialgkgsdaftdtg-sigvmaavsgssvsnfngsg----- 239
```

```
Qy 245 NFGAEYLERRKQRYFVOEGGLKFNNSGKWERDQRPYWKTKYQKYNDOELOKYIEGH 304
Db 240 -----inskefygdk-----ymkqnpksqlyk----- 261
Qy 305 DKSRENLAPOYDITPDPSLKOOSAGNLFKLEYDGFVNKYTAQFRDLNFKIGSRKIIN 364
Db 262 -----mdirpde-----fnsfelsartyenkfrdrdits 290
Qy 365 RNYQFNGLS--LNSYANLNLTAAVNSGROKYPKGSKFTGWLKDKDFETYNNAKLTLDLNN 423
Db 291 ddyiykyhtpfseidfnvcastsrnqkyrdgslytfy-----ktsaqrnsdaldnnt 346
Qy 424 ATFRLPRETELQTTLGFNFYHNEYGKRNFPBELGLFFDGDQDNGLSYILGRFKGDKGLL 483
Db 347 srftv-adndlefmlgsklmrty-----drtihsaagdpkangesi 387
Qy 484 POKSTIVOPAGSQVNTFYFDALKKDI-----YRLNVTNTVGRFGEVYTYGYGSBDEF 539
Db 388 --ennpfapsqqdisalytqkvtgrgiweadfnlnytrnri-----tg----- 429
Qy 540 KRAFGENSPYKKHCNOS--C--GIYEPVLKYYKKRANHNSVISISADFGDYEMPASY 594
Db 430 -----ykpacdsrvicvpggsydidkegg-----fnpsvqlsaqvtwlpqfigy 475
Qy 595 SRTHRMPIQEMYSQIGDSGVHTPALKPERANTWQGFNTYKKGLLKQDDTLGLKLGVYR 654
Db 476 sksmrapniqemffsnsgaamplkperatwqagfnidtrdlveqdalrfkalayr 535
Qy 655 SRIDNYIHN-----VYKWW-----DLNGNIPSWSVSTGLAYTIQHRNF 693
Db 536 srignyfysesylvcsggrkcskpevigngwegisdeysdmnyivnsa----- 584
Qy 694 KDKVHKGFLELNDYDGRPTNLSYAYOKSTOPTNFSDASESPNNASKEDOLKQGYLS 753
Db 585 sdvtak-gfelemdydgafgrlsfsqqtdqptsasth-----fag 628
Qy 754 RVSAALPDYGRLEVTRWLNKLTGLGAMRYFGKSI RATAEERYIDGTNGGNTSNVRQLG 813
Db 629 ditelpkymtldtgvrfddnaltltgiikytkarllspdfegdehtga----- 678
Qy 814 KRSTKQETLARQPLIFDFYAAYPEKNLIFRAEVKNLFDPRYIDPLDAGNDAATQRYYS 873
Db 679 --likq--dlpqbtliidltyeynrltliklsvqnlmordysealnkln----- 725
Qy 874 SFDPKDDEEVTCTNADTLGCKYGGTFSKSVLTNFAFGRTFLITMSYKF 922
Db 726 -mmpglgdethpans-----argrtwifggdfrf 753
```

## RESULT 14

```
AA07476
ID AAY07476 standard; Protein; 915 AA.
XX
AC AAY07476;
XX
DT 17-AUG-1999 (first entry)
DE N.gonorrhoea transferrin binding protein.
XX
KW Transferrin binding protein; iron; outer membrane protein; uptake;
KW Neisseria gonorrhoea; nutrient; growth; Neisseria meningitidis; pathogen;
KW receptor; antibody.
XX
OS Neisseria gonorrhoea.
XX
US 5912336-A.
XX
PN 15-JUN-1999.
XX
PF 23-DEC-1994; 94US-0363124.
XX
PR 23-DEC-1994; 94US-0363124.
```

```
PR 23-AUG-1990; 90US-0572187.
PR 05-NOV-1992; 92US-0973336.
PR 20-SEP-1993; 93US-0124254.
XX
XX (UYNC-) UNIV NORTH CAROLINA.
XX
XX Cornelissen CN, Sparling PF;
XX
XX WPI; 1999-357219/30.
XX
XX N-PSDB; AAX78929.
XX
XX Nucleic acid molecules encoding transferrin binding proteins
XX
XX Disclosure; Fig 1; 38pp; English.
XX
XX This sequence represents the transferrin binding protein from Neisseria
XX gonorrhoea. Transferrin is an iron-regulated, outer membrane protein
XX involved in uptake of iron, an essential nutrient for the growth of
XX N.gonorrhoea and N.meningitidis. A similar sequence from N.meningitidis
XX is shown in AAY07477. N.gonorrhoea and N.meningitidis are two pathogens
XX of the genus Neisseria that are genetically similar, but pathologically
XX different. The growth of these cells can be inhibited by reducing the
XX ability of these cells to take up iron, e.g. by blocking the transferrin
XX receptor function. The transferrin binding proteins can be used to
XX raise antibodies for inhibiting the activity of the transferrin
XX receptor.
XX
XX Sequence 915 AA;
XX
XX Query Match 5.7%; Score 278; DB 20; Length 915;
XX Best Local Similarity 21.4%; Pred. No. 2.2e-11;
XX Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;
XX
Qy 5 FRLPKPCIFLMGVMLYHHSYAEDAGRAGSEAQIQVLEDVHVAK--RVPKDKKVFTHDARA 62
Db 7 frlnilcslmtal---payaeav-qag-qaqekqldtqivkakkqktrrdnevtgigkl 61
Qy 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFRVNTMVDGIT 118
Db 62 vktadtlskeqvidrldtrydpgiavveqrgassg---ysirg-mdknrvsltdvqla 117
Qy 119 Q-TFYSTSTDAG--RAGGSQFQASVDSNFIAGLDVVVKGSFSGSAGINSIAGSANTLRTGL 175
Db 118 qiqsytaqaalgtrtagssgaineleyenvkaveiskgsnsvqsgsalagsvafqtkt 177
Qy 176 VDDVVOGNNTYGLLLK-GLTGTSN--TKGNAMAA-IGARKWL--ESGASVGVLYGHSSRS 229
Db 178 addvigeqrqwigsktaysgknrgltqstalgriqgaallirtqhageirah---- 233
Qy 230 VAQNYRVGGGQGHIGNFGAEYLERRKQRYFVOEGGLKFNNSGKWERDQRPYWKTKWYQ 289
Db 234 -----eaagrgvgsfnrlapvddgskyaayfiveeck-----ngghekckanp----- 276
Qy 290 KYNDPOELQKYECHDK-----SWRENLAPOYDITPDPSLKOOS-----AGNLF--KLEY 339
Db 277 -----kkdvvgedkrqvtsttdytpgnrfla--dplsyerswlfirpgrfrfenkrhy 326
Qy 340 DGVFNKYTAQ---FRDLNFKIGSRKII-----NRNYQFNYSLSNLSYANLN 383
Db 327 iggilletqtqtdrtdntvpafitkavfdanqkqagslrgngkyagrh-----kyggl-- 379
Qy 384 TAAVNSGRQKYPKGSKFTGWLKDKDFETYNNAKLTLDLNNATATFRLPNETELQTTLGFNYF 443
Db 380 ---ftsgennapvgaey-gtgvfyd-----ethtkrsrygleyv 413
Qy 444 HNEYGKRNFPPEELGLFFD--GPDQDNGL-----YSYIGRFRFGDK 480
Db 414 ytnadktdwadyarlsydrqgigldnhfqthcsadgsdkycrpsadkpfisy---yksdr 470
Qy 481 GLLPQKSTIVOPAGSQVNTFYFDALKKDIYRLNVTNTVGY-RFGG-----EYTCYYG 534
Db 471 viygeshklilqaakksfdtakl-----rhnlsvn-lgydrfgsnlrhgdqy---yyq 518
```



QY 535 SDDEFKRAFGENSTYKKHCNQSGGIPEVLKYYGKKRA-----NNHSVISAD----- 583  
 Db 519 san---rayslktpp-----qngkktspngreknpywvsigrgnvvtr 559  
 QY 584 ----FGD-----YMPFA-----SYSRTH----- 598  
 Db 560 qiclfgnntydctprsingksyyaavrdnrlgrwadvgaglrlydrsthsddgsvstg 619  
 QY 599 -----RMPNIQEMYFSQIGDSGVHTALKPERANTWQ 629  
 Db 620 thrtlswnagivlkpadwldltyrtstgfrlpsfaemgywrsgdikavkidpeksfne 679  
 QY 630 FG-----FNTYKKGLLKQDDTLGLKLVGRSRIDN-----YIH--- 662  
 Db 680 agivfkdgdfngleaswfnayrdlrv-----gyeaqikdgkeqvgkgnpaylnaqs 730  
 QY 663 -----NVYQK-W---WDLGNIP-SWVSSTGLAYT-IQHRNFKDKVHKHGFLELYND 709  
 Db 731 aritginilgkidwngvd---klpegwyst--faynrvrdrdkkradrtldigshl-fd 784  
 QY 710 YGRFTNLSTAYQKSTOPTNFSDASESPNNASKEDQLKQGYGLSRVSALEPRDYGRLEVGT 769  
 Db 785 -----aidpsryv-----vgsydydpegkwgvingmit-----yskakeit 819  
 QY 770 RWLGNKLTGGAMRYFCKSIRATAERYIDGTNGGNTSNVRQLGKRSIKQETELARQPLI 829  
 Db 820 elglsrallng-----nsrntkatarrtrpwyi 847  
 QY 830 FDEYAAYEPKKNLIFRAEVKNLFDPRYI 857  
 Db 848 vdvsgyytvkhhftlravynllnhryv 875

RESULT 15  
 AAY51770  
 ID AAY51770 standard; Protein; 915 AA.  
 XX AAY51770;  
 XX  
 DT 13-JUN-2000 (first entry)  
 XX N. gonorrhoeae strain FA19 Tbp1 protein.  
 DE  
 XX  
 KW Transferrin receptor; Tbp1; Tbp2; immunogenic; antibacterial; vaccine;  
 KW diagnosis.  
 XX  
 OS Neisseria gonorrhoeae.  
 XX  
 XX US6015688-A.  
 XX  
 PD 18-JAN-2000.  
 XX  
 XX 07-JUN-1995; 9505-0483577.  
 XX  
 PR 08-NOV-1994; 9405-0337483.  
 PR 08-NOV-1993; 9305-0148968.  
 PR 29-DEC-1993; 9305-0175116.  
 XX  
 XX (CONN-) CONNAUGHT LAB LTD.  
 XX  
 PI Loosmore S, Harkness R, Schryvers A, Gray-Owen S, Yang Y;  
 PI Murlin A, Klein M, Chong P;  
 XX  
 XX WPI; 2000-181144/16.  
 XX  
 XX New nucleic acid encoding truncated transferrin receptor, useful for  
 PT diagnosis, treatment and prevention of bacterial infections,  
 PT particularly by Haemophilus -  
 XX  
 XX Example 6; Column 163-168; 281pp; English.  
 XX  
 XX This invention describes a novel isolated and purified nucleic acid (I)

CC encoding an immunogenic, C-terminally truncated analog of one of the  
 CC transferrin receptor proteins Tbp1 or Tbp2 of Haemophilus influenzae  
 CC which has antibacterial activity. (I) are used for recombinant  
 CC production of truncated Tbp; as probes and primers for detecting, and  
 CC diagnosing infection by, Haemophilus, also for isolating similar  
 CC sequences from other bacteria; as immunogens for vaccinating against  
 CC infections caused by bacteria that produce transferrin receptors, e.g.  
 CC Haemophilus, Neisseria or Branhamella. The truncated proteins are useful  
 CC as immunogens (as above); for diagnosing infection (as antigens in  
 CC immunoassays) and for raising antibodies, used for diagnosis of  
 CC infections or for passive immunization. This sequence represents the  
 CC transferrin receptor protein Tbp1 isolated from Neisseria gonorrhoeae  
 CC strain FA19.  
 XX  
 XX Sequence 915 AA;

Query Match 5.7%; Score 278; DB 21; Length 915;  
 Best Local Similarity 21.4%; Pred. No. 2.2e-11;  
 Matches 224; Conservative 127; Mismatches 333; Indels 374; Gaps 58;

QY 5 FRLKPICFYLMGVMLYHHSYAEDAGRASEAQIQVLEDVHVYKAK--RVPDKKVFDTDARA 62  
 Db 7 frlnilcslmtal---payaenv-qag-qakeqldtiqvakkkqktrrdnevtgikl 61  
 QY 63 VSTRQDIFKSE-NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGRVNWTVDGIT 118  
 Db 62 vktadtiskeqvlrdltdrtdpdlavveqgrgassg---ysirg-mdknrvstvdgla 117  
 QY 119 Q-TFYSTSTDAG--RAGSSQFGASVDSNFTAGLDVVKVSGSAGINSLGASAMRLTLG 175  
 Db 118 qiqsytaqaalgtrtagssgaieieyenvkaveiskgsnveqsgalagsvafqkt 177  
 QY 176 VDDYVQGNNTYGLLLK-GLTGTNS--TKGNAMAA-IGARKWL--ESGASVGLYGHRSRS 229  
 Db 178 advigegrwgiqsktaysgknrgltqslalagriggaeallrtgrhageirah---- 233  
 QY 230 VAQNYRVGGGQHIGNFGAEYLERRKQRYFVOEGGLKFNSSGKWERDFQRPWKTWYQ 289  
 Db 234 -----eaagrgvgsfnrlapvddgskyaifiveeck-----ngghekkcaup----- 276  
 QY 290 KYNDPQELQKYIEGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AGNLF--KLEY 339  
 Db 277 -----kkdvgedkrqvtstvdtygpnrfia--dplesyesrswlfrpfrfenkrhy 326  
 QY 340 DGVFNKYTAQ---PRDLNWKIGSRKII-----NRNYOFNGLSLNSYANLNL 383  
 Db 327 iggilerqtgtfdrtdmtvpaflltkavfdanqkqagslrngkkyagnh-----kygg1-- 379  
 QY 384 TAAYNSGRQYKPKGSKFTGWLKLDKFETYNNNAKILDLNNTATFRLPRETELOTTLGFNYF 443  
 Db 380 ---ftsgennapvgaey-gtgvfyd-----ethckrygleyv 413  
 QY 444 HNEYGNRFPEELGLFFD--GPDODNGL-----YSLGRFGKDGK 480  
 Db 414 ytnadkdtwadyarlsydrdgigldhffqthcsadsdkycrpsadkpfisy---yksdr 470  
 QY 481 GLLPQKSTIVQAGSOYFNTFYDALKKDIYRLNYSTNTVGY-RFGG-----EYTGYYG 534  
 Db 471 viygeshkliaqaafkksfdtaki-----rhnlsvn-lgydrfgsnlrhqdq--yyq 518  
 QY 535 SDDEFKRAFGENSTYKKHCNQSGGIPEVLKYYGKKRA-----NNHSVISAD----- 583  
 Db 519 san---rayslktpp-----qngkktspngreknpywvsigrgnvvtr 559  
 QY 584 ----FGD-----YMPFA-----SYSRTH----- 598  
 Db 560 qiclfgnntydctprsingksyyaavrdnrlgrwadvgaglrlydrsthsddgsvstg 619  
 QY 599 -----RMPNIQEMYFSQIGDSGVHTALKPERANTWQ 629  
 Db 620 thrtlswnagivlkpadwldltyrtstgfrlpsfaemgywrsgdikavkidpeksfne 679

QY	630	FG-----FNYKKGLLKQDDTLGKLKLVGRSRIDN-----YIH----	662
		: : :   : : :   : : :   : : :   : : :   : : :	
Db	680	agivkgdfgnleasvnnayrdilvr-----gyeaikdgkeqvgknpaynsqs	730
QY	663	-----NVYGK--W---WDLNGNIP-SWVSSTGLAYT-IQHNRFNKDVKHKHCFELELYND	709
		:       :       :       :       :       :       :	
Db	731	aritginiigkidwngwd--klpegwyst--faynrvrirdlkkdrfdtighl-fd	784
QY	710	YGRFFTNLSVAYOKSTOPNFSDASESPNNASKEDQLKGQYGLSRVSLPALPDYRGLEYGT	769
		:   :   :   :   :   :   :   :   :   :   :   :   :	
Db	785	-----aiqpsryv-----vgsygdpegkwgvngmlt----yskakeit	819
QY	770	RWLGNKLTLLGGAMRYFGKSIRATABERYIDGTNGGWNTSVNLORGKRISIKQTETLARQPLI	829
		: :     :   :   :   :   :   :   :   :   :   :   :	
Db	820	ellgrstallng-----nsrnlkatarrtpwyi	847
QY	830	FDPYAAYEPKKNLIIFRAEVKNLFDRIYI	857
		:   :   :   :   :   :   :   :   :   :   :   :   :	
Db	848	vdsvytytvtkkhftiragvynllnhyr	875

Search completed: July 24, 2002, 08:53:15  
Job time: 517 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:45:49 ; Search time 42.33 Seconds  
(without alignments)  
2092.945 Million cell updates/sec

Title: US-09-762-926-4  
Perfect score: 4904  
Sequence: 1 MRSSFLKPCIFLYMGVMLY.....SVLTNFARGRTFLITMSYKF 922

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 283138 seqs, 96089334 residues

Total number of hits satisfying chosen parameters: 283138

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : PIR 71: \*  
1: pir1: \*  
2: pir2: \*  
3: pir3: \*  
4: pir4: \*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match %	Length	ID	Description
1	4840	98.7	922	G81865	probable outer mem
2	4768.5	97.2	921	E81076	TonB-dependent rec
3	2514.5	51.3	913	G64110	hypothetical prote
4	1085	22.1	755	A81436	probable outer mem
5	425	8.7	891	H83218	heme acquisition p
6	435	8.7	889	C83035	hypothetical prote
7	414	8.4	830	AH0477	probable TonB depe
8	360.5	7.4	852	C98310	hasR protein (Y089
9	360.5	7.4	923	AH2972	heme receptor hasR
10	308	6.3	784	D82437	TonB receptor-rela
11	278	5.7	915	A43335	transferrin-bindin
12	269.5	5.5	908	JN0819	transferrin-bindin
13	267.5	5.5	910	C81832	transferrin-bindin
14	266	5.4	698	E82443	heme transport pro
15	266	5.4	744	B64049	outer membrane pro
16	262	5.3	766	D97634	probable outer mem
17	262	5.3	766	AF2857	outer membrane hem
18	259.5	5.3	791	F81056	hemoglobin recepto
19	257	5.2	911	JN0821	transferrin-bindin
20	256	5.2	915	F81196	transferrin-bindin
21	243.5	5.0	940	S49087	lactoferrin bindin
22	243	5.0	943	C81070	lactoferrin-bindin
23	242	4.9	912	C64107	transferrin-bindin
24	237.5	4.8	792	S61335	hemoglobin recepto
25	233	4.8	877	AC2211	heme transport pro
26	233	4.8	912	S70901	transferrin-bindin
27	228.5	4.7	911	S70911	transferrin-bindin
28	226.5	4.6	944	C81798	lactoferrin bindin
29	225.5	4.6	843	A87275	TonB-dependent rec

30	223	4.5	764	2	H83055	probable outer mem
31	222	4.5	914	2	S70906	transferrin-bindin
32	218.5	4.5	723	2	C64058	outer membrane pro
33	212	4.3	953	2	B64083	hemoglobin-binding
34	211.5	4.3	725	2	A57148	outer membrane pro
35	210.5	4.3	931	2	S66574	transferrin-bindin
36	205.5	4.2	1084	2	B64088	hemoglobin-binding
37	204.5	4.2	810	2	A81965	hemoglobin-haptog
38	203.5	4.1	720	2	S58133	Fe-regulated prote
39	195	4.0	2817	2	B97033	uncharacterized pr
40	191.5	3.9	851	2	A83484	probable heme util
41	190.5	3.9	660	2	D91176	heme utilization/t
42	190.5	3.9	660	2	E86022	outer membrane hem
43	189.5	3.9	713	2	A56268	Fe-regulated prote
44	185	3.8	714	2	F81962	probable iron-regu
45	185	3.8	795	2	C83626	probable TonB-depe

ALIGNMENTS

RESULT 1  
G81865  
probable outer membrane substrate binding protein NMA1700 [imported] - Neisseria meni  
C:Species: Neisseria meningitidis  
C:Date: 05-May-2000 #sequence\_revision 05-May-2000 #text\_change 02-Feb-2001  
C:Accession: G81865  
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Mo  
; Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandre  
Nature 404, 502-506, 2000  
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491  
A:Reference number: A81775; MUID:20222556  
A:Accession: G81865  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-922 <PAR>  
A:CROSS-references: GB:AL162756; GB:AL157959; MID:g7380091; PIDN:CAB84928.1; PID:g738  
A:Experimental source: serogroup A, strain 22491  
C:Genetics:  
A:Gene: NMA1700

Query Match	98.7%	Score	4840	DB	2	Length	922
Best Local Similarity	98.7%	Pred. No.	8.4e-299				
Matches	910	Conservative	3	Mismatches	9	Indels	0
Gaps	0						
QY	1	MRSSFLKPCIFLYMGVMLYHHSVAEDAGRAGSEAQIOVLEDVHVAKRVPKDKKVF	60				
Db	1	MRSSFLKPCIFLYMGVMLYHHSVAEDAGRAGSEAQIOVLEDVHVAKRVPKDKKVF	60				
QY	61	RAVSTRODIFKSSSENLDNIVRSIPGAF	120				
Db	61	RAVSTRODIFKSSSENLDNIVRSIPGAF	120				
QY	121	FYSTSTDAGRAGSSQFGASVDSNF	180				
Db	121	FYSTSTDAGRAGSSQFGASVDSNF	180				
QY	181	QGNNTYGLLLKGLTGTNSTKGNMAA	240				
Db	181	QGNNTYGLLLKGLTGTNSTKGNMAA	240				
QY	241	QHIGNFGAEYLERRKQRYFVQEGGLKFN	300				
Db	241	QHIGNFGAEYLERRKQRYFVQEGGLKFN	300				
QY	301	IEGHDKSWRENLAPOQDITPIDFSSLKQ	360				
Db	301	IEGHDKSWRENLAPOQDITPIDFSSLKQ	360				
QY	361	KIINRNQFNGVLSLSYANLNLTAAYNS	420				
Db	361	KIINRNQFNGVLSLSYANLNLTAAYNS	420				

QY 421 NNTATRLPRETELOTTLGFNHNEYGNKRFPEELGLFDFGDDQDNGLSYLGRFKGDK 480  
Db 421 NNTSTRLPRETELOTTLGFNHNEYGNKRFPEELGLFDFGDDQDNGLSYLGRFKGDK 480  
QY 481 GLLPQKSTIVOPAGSQYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFK 540  
Db 481 GLLPQKSTIVOPAGSQYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFK 540  
QY 541 RAFGENSPYKHKCNQSCGIEYEPVKKYKGRANNSHVSISADFGDYFMPFASYSRTHRM 600  
Db 541 RAFGENSPYKHKCNQSCGIEYEPVKKYKGRANNSHVSISADFGDYFMPFASYSRTHRM 600  
QY 601 PNTQEMTFYSGISGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKLVGYSRSDNY 660  
Db 601 PNTQEMTFYSGISGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKLVGYSRSDNY 660  
QY 661 IHNVYKQWDLNGNIPSWSTGLAYTIOHNFKDKVHKHGFELNNDYGRFETNLSYA 720  
Db 661 IHNVYKQWDLNGNIPSWSTGLAYTIOHNFKDKVHKHGFELNNDYGRFETNLSYA 720  
QY 721 YOKSTQPTNFSADSESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTGRLGNKLTGG 780  
Db 721 YOKSTQPTNFSADSESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTGRLGNKLTGG 780  
QY 781 AMRYFGKSIRATAEERYIDGTNGTNSNROLGKRSIKQETLARQPLIFDFYAAEYPPK 840  
Db 781 AMRYFGKSIRATAEERYIDGTNGTNSNROLGKRSIKQETLARQPLIFDFYAAEYPPK 840  
QY 841 NLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDEEVTCKNADKTLGNGKYGGT 900  
Db 841 NLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDEEVTCKNADKTLGNGKYGGT 900  
QY 901 SKSVLTNFAARGRTFLITMSYKF 922  
Db 901 SKSVLTNFAARGRTFLITMSYKF 922  
RESULT 2  
E81076  
TonB-dependent receptor NMB1497 [imported] - Neisseria meningitidis (strain MC58 serogroup B, strain MC58)  
C:Species: Neisseria meningitidis  
C:Date: 31-Mar-2000 #sequence\_revision 31-Mar-2000 #text\_change 19-Jan-2001  
C:Accession: E81076  
R:Gottlieb, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.; Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.; Li, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.  
Science 287, 1809-1815, 2000.  
A:Authors: Grandt, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Venturi, A.; Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.  
A:Reference number: A81000; MUID:20175755  
A:Accession: E81076  
A>Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-921 <TET>  
A:Cross-references: GB:AE002499; GB:AE002098; MID:g7226737; PIDN:AAF41853.1; PID:g7226737  
A:Experimental source: serogroup B, strain MC58  
C:Genetics:  
A:Gene: NMB1497

Query Match 97.2%; Score 4768.5; DB 2; Length 921;  
Best Local Similarity 97.7%; Pred. No. 2.9e-294;  
Matches 901; Conservative 8; Mismatches 12; Indels 1; Gaps 1;  
QY 1 MRSSFLKPCIFLGMVLMYHHSVAEDAGRAGSEAOIQVLEDVHVKAKRYPKDKKVFDTA 60  
Db 1 MRSSFLKPCIFLGMVLMYHHSVAEDAGRAGSEAOIQVLEDVHVKAKRYPKDKKVFDTA 60  
QY 61 RAVSTRQDIFKSSNLDNIVRSIPGAFQDQKSSGIVSLNIRGDSGFRVNTMWDTGTOT 120  
Db 61 RAVSTRQDIFKSSNLDNIVRSIPGAFQDQKSSGIVSLNIRGDSGFRVNTMWDTGTOT 120

QY 121 FYSTSDAGRAGSSQFGASVDSNFTAGLDVVKGSPSGSAGINSLAGSANLRLTGVDDVV 180  
Db 121 FYSTSDAGRAGSSQFGASVDSNFTAGLDVVKGSPSGSAGINSLAGSANLRLTGVDDVV 180  
QY 181 QGNNTYGLLLKGLTGTNSTKGNMAAATGARKWLESASVGLYGHRRSVAQNVYRGGGG 240  
Db 181 QGNNTYGLLLKGLTGTNSTKGNMAAATGARKWLESASVGLYGHRRSVAQNVYRGGGG 240  
QY 241 OHIGNFGAEYLERRKQRYFYQEGGLKFNNSGKWERDFORPYWKTWKYQKYNPQELQKY 300  
Db 241 OHIGNFGAEYLERRKQRYFYQEGGLKFNNSGKWERDLQROQWKYKYNKN--QELQKY 299  
QY 301 IEHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEVDGVFNKNTAOPRLNTKIGSR 360  
Db 301 IEHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEVDGVFNKNTAOPRLNTKIGSR 359  
QY 361 KIINRNYQFNYSLSANLANLTAAVNSGRQYKPKSGKFTGMLLKDFETYNNAKILDL 420  
Db 361 KIINRNYQFNYSLSANLANLTAAVNSGRQYKPKSGKFTGMLLKDFETYNNAKILDL 419  
QY 421 NNTATRLPRETELOTTLGFNHNEYGNKRFPEELGLFDFGDDQDNGLSYLGRFKGDK 480  
Db 421 NNTATRLPRETELOTTLGFNHNEYGNKRFPEELGLFDFGDDQDNGLSYLGRFKGDK 479  
QY 481 GLLPQKSTIVOPAGSQYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFK 540  
Db 481 GLLPQKSTIVOPAGSQYFNTFYDAALKKDIYRLNYSNTVGYRFGGEYTGYYGSDDEFK 539  
QY 541 RAFGENSPYKHKCNQSCGIEYEPVKKYKGRANNSHVSISADFGDYFMPFASYSRTHRM 600  
Db 541 RAFGENSPYKHKCNQSCGIEYEPVKKYKGRANNSHVSISADFGDYFMPFASYSRTHRM 599  
QY 601 PNTQEMTFYSGISGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKLVGYSRSDNY 660  
Db 601 PNTQEMTFYSGISGVHTALKPERANTWQFGFNTYKGLLKQDDTLGLKLVGYSRSDNY 659  
QY 661 IHNVYKQWDLNGNIPSWSTGLAYTIOHNFKDKVHKHGFELNNDYGRFETNLSYA 720  
Db 661 IHNVYKQWDLNGNIPSWSTGLAYTIOHNFKDKVHKHGFELNNDYGRFETNLSYA 719  
QY 721 YOKSTQPTNFSADSESPNNASKEDQLKQ 780  
Db 721 YOKSTQPTNFSADSESPNNASKEDQLKQ 779  
QY 781 AMRYFGKSIRATAEERYIDGTNGT 840  
Db 781 AMRYFGKSIRATAEERYIDGTNGT 839  
QY 841 NLIFRAEVKNLFDRIYDPLDAG 900  
Db 841 NLIFRAEVKNLFDRIYDPLD 899  
QY 901 SKSVLTNFAARGRTFLITMSY 922  
Db 901 SKSVLTNFAARGRTFLITMSY 922  
RESULT 3  
G64110  
hypoetical protein H11217 - Haemophilus influenzae (strain Rd KW20)  
C:Species: Haemophilus influenzae  
C:Date: 18-Aug-1995 #sequence\_revision 18-Aug-1995 #text\_change 08-Oct-1999  
C:Accession: G64110  
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage  
R:Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman  
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghegan, N.S.M.  
Science 269, 496-512, 1995  
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter  
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.  
A:Reference number: A64000; MUID:95350630  
A:Accession: G64110  
A>Status: nucleic acid sequence not shown; translation not shown  
A:Molecule type: DNA



Qy 725 TQPTNFSDASESPNNASKEDQKQGYGLSRYSALPRDYRGLEVTGRLWGNKLTLLGGAMRY 784  
Db 594 M-----YSRQDTSSTISQTSGLLSFSASKIMELPKDYANVELGFR-LNDKISFGGIAY 648  
Qy 785 FGKSIRATAEERYIDGNGGHTS-NVRQLGKRSTKQTETLARQPLIFDFFAAYEPKKNLI 843  
Db 649 TGKA-----KRVAPNTDDWKNKDPNNPYPKPT---TQDLPKIPIIIVDLWNIEWFNLT 699  
Qy 844 FRAEVKNLFDRYIDPLDAGNDAATQRYSSFDPKDKDEEVTGNADKTLGCKYGGTSSKS 903  
Db 700 MRAEVQNLFDKNYMDALNAYNSLONQLO-----NG--AGDPIY 736  
Qy 904 VLTNFARGRTFLITMSYKF 922  
Db 737 LFSNSARGRTFIVSFYKY 755  
RESULT 5  
heme acquisition protein HasR PA3408 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: H83218  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; Braham, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lim, S.; Lory, S.; Olson, M.V.  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: H83218  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-891 <STO>  
A:Cross-references: GB:AE004762; GB:AE004091; NID:g9949544; PIDN:AAG06796.1; GSPDB:GN001  
C:Experimental source: strain PA01  
C:Genetics:  
A:Gene: hasR; PA3408

Query Match 8.7%; Score 425; DB 2; Length 891;  
Best Local Similarity 22.7%; Pred. No. 6.5e-19;  
Matches 215; Conservative 114; Mismatches 382; Indels 238; Gaps 37;  
Qy 30 RAGSEAOIQVLEDVHVAKRVPKDKVFTDARAVST--RQDIEKS-SENLDNIVRSIPGA 86  
Db 125 RVSQDILLVQMSPSV-ISAAR--PDDVVVQTPHSVSIGREQIERNPPRHAADMLEETPGV 181  
Qy 87 FTQDQKSSGIVSLNIRGDSFGFRVNTWVDGITQTFYSTDAGRAGSSQFGA-SVDSNF 145  
Db 182 YSSVSQDDPGLSVNIRGTDYGRYNNMSVDGMRQNYQ-----QSGHQQRNGTLVYDPEL 234  
Qy 146 IAGLDVYKGSFSGSAGINSLAGSANLRTGLVDVVOGNNYTG---LLKGLTG-TNSTK- 200  
Db 235 LSEVVIDKGASSAMGAGVIGGIANTFLRDLVRLGKGVGRVLTSLGGDANGTHF 294  
Qy 201 -GNMAAIGARKWLSEASGVLYGHSSRVSQAQRYRGGGQHQHTGFEAGYLERRKORYF 259  
Db 295 IGSAFAAIGTEVW-----DMLVAASERHLG-DYDPGTKGS-IG-----ELRTCAWF 338  
Qy 260 VQEGGLKFNNSGKWERDFORPYWKTWKYQKNDPQELQKYIEGHGKSWRENLAPOVDIT 319  
Db 339 NPEAGORVK-----HSPVAYSQVY-----MRSRLAKLGLVAL 369  
Qy 320 PIDFSSLLKQOSAGNLFKLEVDGVFNKYTAQFRDLNTKIGSRKLIINRYQFNGLSL-NSY 378  
Db 370 PQDQ---RLQFSYLTQVSSYDDA-NMLNTENQALWEKLGSSDVAQNAFIDYGVAPNPL 425  
Qy 379 ANLNLTAAY--NSGRQKYPKSGFTGWLGLKDFETYNNAKILDLNNTATFELPRETELQT 436  
Db 426 VDFRAKLYYVDNRNRQOTLQRTGTPGYSIYYQTDYTG----AQAOQNTSTALDLSFLRA 481  
Qy 437 TLGNFYHNEYGKNR-----PPEELGLFFDGDQDQNGLYSYLGRFKGDKGLLQP 485

Db 482 NYGLEFFDYKVRPDSQPRASTSAVGPPAAEGM---TPKGDRLAGLSLFARLDYD----- 532  
Qy 486 KSTIVQAGSQYNTFFYFDAALKDIYRL-----NYSTNTVGYRFGGEYTYGYSGDSDEF 539  
Db 533 -----YDDWLNINAGRLYDRYRLRGDYGFNARTILGTRTQDPLQFAVD--- 578  
Qy 540 KRAFGENSPTYKKHCNCSOGIYEPVKYKGRANNHVSISADFG-DYFMPFASYSRTH 598  
Db 579 -REEGRFSPTF-----GLSVKPGVDWLQLFATYCKGW 609  
Qy 599 RMENIQEMYSQIDSG-----VHTALKPERANTWQFGNTYKGLLKQDDTLGLKLVG 652  
Db 610 RPPAVTESLITGRPHGGGAENMYFNPFLSPERSKAWGVFNVLKENLWFSDRLGLKWAY 669  
Qy 653 YRSRIDNYIHNVYKQWDLNGNIPSWSSYSLGAVTIQHRNFKDKVHKHGFLELNYDGR 712  
Db 670 FDRVDDFIIFMGGMQ-----PPGYGMAGIGNSAVNNL-DSTRFRGVEYQIDYDAGL 721  
Qy 713 FFTNLSYAYQKS-----TQPTNFSDASESPNNASKEDQKQGYGLSRV--- 755  
Db 722 AYQOLSYYTHMIGSNDFCSTAWLGGVQTQVKSGRRPPVIDMRPDEQANAATHCSAVLGS 781  
Qy 756 -SALPRDYGRLEVGTRWLNKLNILGGMARYFGKSIRATAEERYIDGTNGGNTSNVRQLGK 814  
Db 782 AEHPMPMDRG-----SLTLG--MRFDRRLDYGARARYSEGYSVAGGATVSQAGV 828  
Qy 815 RSIKQETTLARQPLIFDFYAAYPEKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSS 874  
Db 829 YPADWKEY-----TVYDLYGSYRVSDDELTLRLAMENTDRLAYLVPLG----- 870  
Qy 875 FDPKDKDEEVTGNADKTLGCKYGGTSKSVLT-NFARGRTFLITMSYKF 922  
Db 871 -----DVLFTLGRGRTLQGTLEYQF 891  
RESULT 6  
C83035  
hypothetical protein PA4897 [imported] - Pseudomonas aeruginosa (strain PA01)  
C:Species: Pseudomonas aeruginosa  
C:Date: 15-Sep-2000 #sequence\_revision 15-Sep-2000 #text\_change 31-Dec-2000  
C:Accession: C83035  
R:Stover, C.K.; Pham, X.Q.; Erwin, A.L.; Mizoguchi, S.D.; Warrenner, P.; Hickey, M.J.; adman, S.; Yuan, Y.; Brody, L.L.; Coulter, S.N.; Folger, K.R.; Kas, A.; Larbig, K.; Lory, S.; Olson, M.V.  
A:Title: Complete genome sequence of Pseudomonas aeruginosa PA01, an opportunistic pathogen  
A:Reference number: A82950; MUID:20437337  
A:Accession: C83035  
A:Status: preliminary  
A:Molecule type: DNA  
A:Residues: 1-989 <STO>  
A:Cross-references: GB:AE004902; GB:AE004091; NID:g9951162; PIDN:AAG08282.1; GSPDB:GN001  
C:Experimental source: strain PA01  
C:Genetics:  
A:Gene: PA4897  
Query Match 8.7%; Score 425; DB 2; Length 989;  
Best Local Similarity 23.2%; Pred. No. 7.5e-19;  
Matches 230; Conservative 130; Mismatches 382; Indels 250; Gaps 42;  
Qy 30 RAGSEAOIQVLEDVHVAKRVPKDKVFTDARAVSTQDIFKSSSENLD-----NIVR 81  
Db 149 RMSGEAPADLSPPVVSAAELADPKETYT-----APSSVYLSSEDDIDRFRVSVGDLQ 203  
Qy 82 SIGCAFTQDQKSSGIVSLNIRGDSGGFRVNTWVDGITQTFYSTDAGRAGSSQFGASV 141  
Db 204 GIFGVQVDSRNGGALDINIRGIQGSQSRVAVRVDGAQ-----ALDVIYRGYAGTQQRSYI 258  
Qy 142 DSNFIAGLDVVKGSFSGSAGINSLAGSANLRTGLGVDDVVOGNNYTGILLKGLTCTNSTKG 201  
Db 259 DPDLVSSVTVDKGFSPGTRSGAI---GGSEVMRTIGVKDILVDGKDLGVFRFTGWNN---G 312



```
Qy 908 FA--RGRTFLTMSYKF 922
Db 814 FAPSRGRTIOGGFEYKF 830

RESULT 8
C98310
hasR protein (Y08983) [imported] - Agrobacterium tumefaciens (strain C58, Cereon)
C:Species: Agrobacterium tumefaciens
C:Date: 22-Oct-2001 #sequence_revision 22-Oct-2001 #text_change 11-Jan-2002
C:Accession: C98310
R:Goodner, B.; Hinkle, G.; Gattung, S.; Miller, N.; Blanchard, M.; Qurolo, B.; Goldman,
A.; Liu, F.; Wollam, C.; Allinger, M.; Doughty, D.; Scott, C.; Lappas, C.; Markelz, B.;
Science 294, 2323-2328, 2001
A:Title: Genome Sequence of the Plant Pathogen and Biotechnology Agent Agrobacterium tum
A:Reference number: A97359; PMID:11743194
A:Accession: C98310
A>Status: preliminary
A:Molecule type: DNA
A:Residues: 1-852 <KUR>
A:Cross-references: GB:AE007870; PIDN:AAK90005.1; PID:g15159974; GSPDB:GN00170
C:Genetics:
A:Gene: AGR_L_2875
A:Map position: linear chromosome

Query Match 7.4%; Score 360.5; DB 2; Length 852;
Best Local Similarity 21.1%; Pred. No. 7.9e-15;
Matches 201; Conservative 117; Mismatches 332; Indels 303; Gaps 37;

Qy 62 AVSTRQDIFKSS-ENLDNIVRSIPGAFTQDDKSS-GIVSLNIRGDSFGFRVNTWVDGITO 119
Db 111 SVWSREAIQSAGVRNTRDFVNRVSGVYAGGNGSFPTVSPNVRGLQESGRVYVSDIGARQ 170

Qy 120 TF---YSTSTADGAGSSQFGAS-----VDSNFIAGLDVYVKGSGFSAGINSIAGSANL 171
Db 171 NAQRGYST-----GGASIYSANNGQAYVDAAFIRAVEVEKMTSATSGNAGSLGGKVEF 223

Qy 172 RTLGVDVVOGNNYTYGLLLKGLTGTN--STKGNMAAIGARKWLESAGVGLYCHSRRS 229
Db 224 RTVSAADLIPEGANKGAENVNVRSGNGYDFOGSVLAAV-----REPDGPLSEFVAGYS-RT 277

Qy 230 VAQNRVGGGGQHI-----GNFG---AEYLERKQRVFOV--- 261
Db 278 IMDEYKIGTKGEALSTALTMTKDLLGRDGSWTFPFKSEGDFGVDQVTSLSWMHQNDVFQOGAS 337

Qy 262 ---EGGLKFNNSGK--WERDFORPYWTKYQKYNQDPOELQYIEGHDKSWRENLAPO 315
Db 338 TVVDREGVRNDSIVAKLDWDPESELIDFKSSLWLDNNTHLRTHEL-----AART 384

Qy 316 YDI-TPID-----PSSLKOOSAGNLFKLEYDGVFNKYTAQ--FRDLNTKIGSRKIINRY 367
Db 385 YAVETNLDMLGRSFGGSLENTSR--FDTRAGALSINYGAEAFRDIATSVATSIQNP 441

Qy 368 QFNVLGLNSVANL-----NLTAAYNSGROKYPKSGKFTG-----WGLLKDFETYNNAK 416
Db 442 SF-----ASSYTSFSPARRDVASLFLNGELEPADWITLSGGVRYDWSRLKGSATYTSFK 496

Qy 417 ILDLNNTATFRLPRETELQTTFLGNYFNHXYGKNRFPPEELGLFPDQDQNGLYSYLGRF 476
Db 497 -----ESIVTSPVCDLVRNHY----- 513

Qy 477 KDGKLLPQKSTIVOPAGSOYNTFYDAAALKDKDIYRLNYSNTVGYRFGGEYTYGYGSD 536
Db 514 -----TALEYFNQVFLPANLPVNASRYNVLASIW----- 543

Qy 537 DEPKRAFGENSPYKKHCNCSGIEYEPVLKYGKRRANNHSVISADFG----- 585
Db 544 -----PRTSANCMPTGTI-----TTKTPVTEYPSHEVDIDRTYSAWLPSATIEFK 588

Qy 586 --DYFMFPFASYSRTHRPNIQEMFYF--SQIGDS-----GVHTALKPERANTWQGFNTYK 636
Db 589 PVDWFRPVVSVYSQSLRPTILETFEAGARFQDSAGYEPNQSLRAEKATYIEIGANNKF 648
```

```
Qy 637 KGLLKQDDTLGLKLVGYSRIDNYI-----HNVYKGWMDLNGNIPSWSVSTGLATIQ 689
Db 649 DGVLLDDDTLRKMAAFREVKVDYIALGYLVTDQVFDRTY-----TSFVNLDDTTY--- 699

Qy 690 HRNFKDKVHKHGFPELELNYYDGRFETNLVSAYQKSTOPTNFSDESASPNNASKEDQLKQG 749
Db 700 -----MRGLENGYDARSWIGGSATVLKTEWPEKTQVFSNSTTTTSGE----- 744

Qy 750 YGLSRVSALPRDYGRLEVGTRFWLGNKLTGLGAMRYFGKSIIRATAEERYIDGTNGGNTSNV 809
Db 745 -----IVAWPGD-----VAPKM-----KITLDGGMRFDEKFSGLGARLNHVTPTQSRITLDT 791

Qy 810 RQLGKRISIKQETIARQPLIFDFYAAVEPKKNLIFRAEVKNLFRDRYIDPLDAGNDAATQ 869
Db 792 GNL--REITDPYTTV-----DLVGSYAFNDKATLRFAVNNLTDRKYIPRASA----- 836

Qy 870 RYSSFPDKDEEVTCTNADKTLGNGYGTSTKSVLNFARFGRFTLTMSYKF 922
Db 837 --YT-----APGRTFIATMNVKF 852
```

## RESULT 9

AH2972

heme receptor hasR [imported] - Agrobacterium tumefaciens (strain C58, Dupont)  
C:Species: Agrobacterium tumefaciens  
C:Date: 11-Jan-2002 #sequence\_revision 11-Jan-2002 #text\_change 11-Jan-2002  
C:Accession: AH2972  
R:Wood, D.W.; Setubal, J.C.; Kaul, R.; Monks, D.; Chen, L.; Wood, G.E.; Chen, Y.; Woo  
erage, G.; Gillet, W.; Grant, C.; Guenther, D.; Kutayin, T.; Levy, R.; Li, M.; MCCI  
; Karp, P.; Romero, P.; Zhang, S.  
Science 294, 2317-2323, 2001  
A:Authors: Yoo, H.; Tao, Y.; Biddle, P.; Jung, M.; Krespan, W.; Perry, M.; Gordon-Kam  
ster, E.W.

A:Title: The Genome of the Natural Genetic Engineer Agrobacterium tumefaciens C58.  
A:Reference number: AB2577; PMID:11743193  
A:Accession: AH2972

A&gt;Status: preliminary

A:Molecule type: DNA

A:Residues: 1-923 &lt;KUR&gt;

A:Cross-references: GB:AE008689; PIDN:AAL44198.1; PID:g1774177; GSPDB:GN00187

A:Experimental source: strain C58 (Dupont)

C:Genetics:

A:Gene: hasR

A:Map position: linear chromosome

Query Match 7.4%; Score 360.5; DB 2; Length 923;  
Best Local Similarity 21.1%; Pred. No. 8.4e-15;  
Matches 201; Conservative 117; Mismatches 332; Indels 303; Gaps 37;

```
Qy 62 AVSTRQDIFKSS-ENLDNIVRSIPGAFTQDDKSS-GIVSLNIRGDSFGFRVNTWVDGITO 119
Db 182 SVWSREAIQSAGVRNTRDFVNRVSGVYAGGNGSFPTVSPNVRGLQESGRVYVSDIGARQ 241

Qy 120 TF---YSTSTADGAGSSQFGAS-----VDSNFIAGLDVYVKGSGFSAGINSIAGSANL 171
Db 242 NAQRGYST-----GGASIYSANNGQAYVDAAFIRAVEVEKMTSATSGNAGSLGGKVEF 294

Qy 172 RTLGVDVVOGNNYTYGLLLKGLTGTN--STKGNMAAIGARKWLESAGVGLYCHSRRS 229
Db 295 RTVSAADLIPEGANKGAENVNVRSGNGYDFOGSVLAAV-----REPDGPLSEFVAGYS-RT 348

Qy 230 VAQNRVGGGGQHI-----GNFG---AEYLERKQRVFOV--- 261
Db 349 IMDEYKIGTKGEALSTALTMTKDLLGRDGSWTFPFKSEGDFGVDQVTSLSWMHQNDVFQOGAS 408

Qy 262 ---EGGLKFNNSGK--WERDFORPYWTKYQKYNQDPOELQYIEGHDKSWRENLAPO 315
Db 409 TVVDREGVRNDSIVAKLDWDPESELIDFKSSLWLDNNTHLRTHEL-----AART 455

Qy 316 YDI-TPID-----PSSLKOOSAGNLFKLEYDGVFNKYTAQ--FRDLNTKIGSRKIINRY 367
```





Db 456 YAVETNLDMLRSGGLENTSR---FDTRAGALSLNYGAEAPRDIAISVATSAIAQNP 512

Qy 368 QFNVGLSLSYANL-----NLTAAYNSGRQYKPGSKFTG-----WGLLKDFETYNNAK 416

Db 513 SF-----ASSYTSFSPAGRRDVAFLNGEPAADWITLSSGGVRYDMSRLKGSATYYSFK 567

Qy 417 ILDLNNTATFRLPRETELQTLGNYEHNEYGNRPPEELGLFDFGPDQDNGLSYLGFR 476

Db 568 -----ESTVTSVPCDLVRNH----- 584

Qy 477 KGDKGLELPQKSTIVQAGSOVFNTEYDAALKKDIYRLNSTNTVGYRFGEGEYTYGSD 536

Db 585 -----TALSYFNQVFLPANLPWASRYNVFLASIM----- 614

Qy 537 DEFKRAFGSPYKHKCNOSCGIYEPVLKYGKKRANHSVSISADFG----- 585

Db 615 -----PRTSANCMPTGT-----TTKTPVTEYPSHEVDIDRTYSAWLPSATIEFK 659

Qy 586 --DYEMFPASYSRTHRNPTQEMFY--SOIGDS-----GVHTALKPERANTWQGFNTYK 636

Db 660 PVDWFRPYVSYSLRPPTILETFAGARPGDSAGYEPAPNQSRLRAEKATTYEIGANMSF 719

Qy 637 KGLLKQDDTLGLKLVGYSRIDNYI-----HNVYKGMWDLNAGNIPSWSVSTGLAYTIQ 689

Db 720 DGVLLDDTLRIKMAAFRRVKDYIALGYLVTQVFDRTY-----TSFVNLDTTY--- 770

Qy 690 HRNFKDKVHKHGFLELNYDYGRFTWLSYAYOKSTOPTNFSDAESPNNASKEDQLKQG 749

Db 771 -----MRGLENGYDARSFWIGSATVLKTEWPEKTQVFSNNTTTTSGE----- 815

Qy 750 YGLSRVSLPRDYGRLVGRWGLGNKLTGAMRYFKSIRATAEERYIDGTNGGNTSNV 809

Db 816 -----IVANFGD-----VAPKM-----KUTLGGMRFFDEKFSGLARLNHVPTQSRILTDE 862

Qy 810 ROLGKRRIKOTETLARQPLIFDYAAYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQ 869

Db 863 GNL--REITDPYTTV-----DLYGSYAENDKATLRFVAVNNLTDKRYIPAASA----- 907

Qy 870 RYSSFPDPKDKDEVTCAADKTLNKGKGTGTSVLTFNFAFGRTFLITMSYKF 922

Db 908 --YT-----APGRFTFIATMNVKF 923

RESULT 10

D82437

TonB receptor-related protein VCA0625 [imported] - Vibrio cholerae (strain N16961 serog

C:Species: Vibrio cholerae

C>Date: 18-Aug-2000 #sequence\_revision 20-Aug-2000 #text\_change 02-Feb-2001

C:Accession: D82437

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwinn, M.L.; Dodson, R.J.;

Chardson, D.; Ermolaeva, M.D.; Vamathevan, J.; Bass, S.; Qin, H.; Dragoi, I.; Sellers, H.

l, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.

Nature 406, 477-483, 2000

A:Title: DNA sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833

A:Accession: D82437

A>Status: preliminary

A:Molecule type: DNA

A:Residues: 1-784 <HEI>

A:Cross-references: GB:AE004392; GB:AE003853; NID:g9658031; PIDN:AAF96526.1; GSPDB:GN001

A:Experimental source: serogroup O1; strain N16961; biotype El Tor

C:Genetics:

A:Gene: VCA0625

A:Map position: 2

Query Match 6.3%; Score 308; DB 2; Length 784;

Best Local Similarity 21.2%; Pred. No. 1.4e-11;

Matches 202; Conservative 115; Mismatches 314; Indels 320; Gaps 46;

Qy 6 RLKPCIFLGMVLMYHHSYAEDAGRAGSEAOIQVLEDVHVHAKRVKPKKVFVDARAVST 65

Db 36 QLKLPLFTLLPVLLSSVVQAQE-----NTEQAVDETIVH-----GQSILTDQR---T 79

Qy 66 RQDIFKSSSELDNIVRSIPGAFTQQDKSSGIVS-----LNTRGDSGGRV 111

Db 80 RSLDLK-----VRGIANA-----DIFSGITSVQSNMHNHEAGALDIGIRGQGEGRVP 127

Qy 112 TWVDGITOTFYSTDAGRAGSSQFGASVDSNFIAGLDVVKG-----FSGSAGINS 164

Db 128 IFIDGLQ---SHTSRGYOGVSDR---TYIDTLSSLTVNKGATIESSPYAGAVG--- 179

Qy 165 LAGSANTRLTGVDVVQGNNTYGLLLKGLTGTNSTKGNMAAICARKWLESASGVLYG 224

Db 180 --GVNATTLIGINDIKDDQAGVWLKARANHNRTDV-----SGDYSEGO-----YA 227

Qy 225 HSRSSVAONTRVG-----GGQHGIGNFGA--EYLERRKQRYFVQEGGLKFNNSGKWERDF 278

Db 228 LDERGEHSAPFKHGLSLMLGLGYQAESFNTVLAYSKRSGKNHF-----AGK----- 271

Qy 279 QRPYWKTKWYKNDP-----QELQKYIEGHDKSWRENLAPOYDITPIDPSSLKQQSAGN 333

Db 272 -----KGTEEQEPVVGQGVVNTSFESD--SW----- 298

Qy 334 LFKLEYDGVFNKYTAQFRDLNTRIGSRKIIINRNYQFNGLSLNAYANLNTAAYNSGRQK 393

Db 299 LFKLASD---TGTAHNADFNRRHQAQKACEVLWAYWYKSSDEWEGN-----PYPDCKDR 349

Qy 394 YPKGSFTGHLGLKDFETYNNAKILDLNNTATFRLPRETELQTLGTFNYFHN-----EY 447

Db 350 MPQ-----WGL-----GTAKV-----NTYSANYYYQDPHPWLNL 378

Qy 448 GKRRFPEELGLFFDGDODNGLYSLGRFKGDKGLLPKSTIVQAGSOVFNTEFYFDAA-- 506

Db 379 NANFWTEADL-----AQYGLWA-LG-----TNAEQFHHYHNDRSG 415

Qy 507 -----LKKDIYRLNSTNTVGYRFGEGYTYGYGSDDEFKRAFGENSTYKHKHNSC 558

Db 416 LSLTNETLLTQWPVRLNYGLAQONERLSPEEDG---QTRFTKT----- 455

Qy 559 GIYEPVLKYGKKRANHSVSIADFG-----DY----- 587

Db 456 -----VTSRHGKRTAQNLFANADIDYSLRVQLGNLHNNAKSTDYQTKQOLDYKEKIDLL 510

Qy 588 -----FMP-----FASYSRTHRNPTQEMFYFS-QIGDSGVHTALKPERANTWQGFNTYK 636

Db 511 SEETVALTPSTQLFLKSSRTYRMPSLYETTLNSNEVFSYNPNPIKPEQANNNEVGQVMA 570

Qy 637 KGLLKQDDTLGLKLVGYSRIDNYIHNVYKGMWDLNAGNIPSWSVSTGLAYTIQHRNEKDK 696

Db 571 SNSVLQDDRLNLSVSYFRNSIKDFISG--GRLAKTPG-MSEW--QANFTFT---NT-DK 620

Qy 697 VHKHGELELNYDYGREFTNLVAYQKSTOPTNFSDAESPNNASKEDQLKQVGLSRVS 756

Db 621 LQLSGMELGAHYQYAWLYTHFAATLYSETKICSVQQAQYAESDTC---NSLGFAMGLTPTR 678

Qy 757 ALPRDYGRLEVGTRWLGKNTLGLGAMRYF--GKSIRATAEERYIDGTNGGNTSNVRLQGLKR 815

Db 679 IPPQNLYLVNGHKFFNDTLDGSKVKSYSYHSGKS-----NPSDWLAGT----- 720

Qy 816 SIKQETLARQPLI-----FDYFAAYEPKKNLIFRAEVKNLFDRIYDP 859

Db 721 -----AANPILEIPSDYITIDLYSQVELNANTQLFFAINNVTDRYOVRP 763

RESULT 11

A43335

transferrin-binding protein 1 - Neisseria gonorrhoeae

C:Species: Neisseria gonorrhoeae

C>Date: 10-Jun-1993 #sequence\_revision 18-Nov-1994 #text\_change 26-Aug-1999

C:Accession: A43335

R:Corneilissen, C.N.; Biswas, G.D.; Tsai, J.; Paruchuri, D.K.; Thompson, S.A.; Sparlin

J. Bacteriol. 174, 5788-5797, 1992

A:Title: Gonococcal transferrin-binding protein 1 is required for transferrin utiliza

A:Reference number: A43335; MUID:92394880

A:Accession: A43335

Db	327	HYVGAVLERQTOQTFDRMTVPAYFTSEDYVPGSLKLGKY-----SGDNKAERL	376
QY	397	-----GSKFTGWGLLKDFETYNNAKILDLNNTATFRLPRETELOTTLGFNFYHNEYGKNR	451
Db	377	FVQGEGSTLQIG-----YGTGVFYD-----ERHTKNRYGVVEYVYHNADKOT	418
QY	452	FPEELGLFFD--GPDODNGL-----YSYLGREFKCDKGLLPQKST	488
Db	419	WADYAKLSYDRQGDIDLNRLQOQTHCHSDGSKNCRPDGNKPYSF---YKSDRMIEESRN	475
QY	489	IVQAGSOFYNT-----FYEDAAL-----	507
Db	476	LFQAVEKKAFTAKIRHNLISINLYGDRXSQLSHSDYILQNAQAYDLITPKPPFPNGS	535
QY	508	KKDIYRLNLTNTVG-----YRFGGE-YT-----GYGSDDEFKRAFGENSPTYKK	552
Db	536	KNPYRVSTGKTTVNTSPICRFGNNTYTDCTPRNIGNGYI-----AAVDQNVRLGR	587
QY	553	HCNOSCGI-YEPVLKKYKVKRANHHSVS-----ISADFGDYFMPFASYSRTH-----R	599
Db	588	WADVAGAIRYD-----YRTHSEDKSVSTGTHRNLSWAGVVLKPFTHWDLTYRSTGFR	642
QY	600	MPNIQEMYSQIGDSGVHTALKPERANTWQFG-----FNTYKGLLKQDDT	645
Db	643	LPSEAEMYGWRAGESLKTLDLKEKSFNREAGTVFKGDFCNLEASVENAYRDLI-----	697
QY	646	LGKLVGYSRDN-----YTH-----NVYCK--WDLNGNIPSWSSSTGL	684
Db	698	-----AFGYETRTONGOTSASGDPGYRNAQNAIAGINILKIDWHGVWGLPDLGYST-L	752
QY	685	AY-----TIQHRNPKDKVHKHGFLELNYDY--GRFTNLSYAYOKSTQPT	728
Db	753	AYNRIKVKADIRADRTFVTSYLFDAVQPSRYVLGLGYDHPDGIWGINTFTYSK-----	807
QY	729	NFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLVGTGRWLNKLTGLGAMRYFGKS	788
Db	808	-----AKSVDEL-----LGSQALLNG-----	823
QY	789	IRATAERYIDGTNGNTSNVROLGKRSIKQTTTLARQLIFDFYAAYPEKKNLIFRAEV	848
Db	824	-----NANAKKAASRR-----TRPWVTVDSVGYNIKKHLTLRAGV	859
QY	849	KNLFDRIYI	857
Db	860	YNLNLTRYV	868
RESULT 13			
C81832			
Transferrin-binding protein A NMA2024 [imported] - Neisseria meningitidis (strain 22491			
C:Species: Neisseria meningitidis			
C>Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001			
C:Accession: C81832			
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel			
S.: Holroyd, S.; Jagels, K.; Leather, S.; Moule, S.; Mungall, K.; Quail, M.A.; Rajandream,			
Nature 404, 502-506, 2000			
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.			
A:Reference number: A81775; MUID:20222556			
A:Accession: C81832			
A>Status: preliminary			
A:Molecule type: DNA			
A:Residues: 1-910 <PAR>			
A:Cross-references: GB:AL162757; GB:AL157959; NID:g/380371; PIDN:CAB85243.1; PID:g/38065			
A:Experimental source: serogroup A, strain 22491			
C:Genetics:			
A:Gene: tbpA; NMA2024			
C:Superfamily: bacterial pathogen transferrin-binding protein; tonB-dependent receptor a			
Query Match			
Best Local Similarity 5.5%; Score 267.5; DB 2; Length 910;			
Matches 215; Conservative 124; Mismatches 355; Indels 329; Gaps 51;			

QY	5	FRLKPICFYLMGVMLYHHSYAEDAGRAGSEAQIQVLEDVHVKAK--RVPKDKKVFVDARA	62
Db	7	FRNLILSLMTAL--PVTAEV--QAG-QAQBKQDLTIQVKKAKQKTRDNEVTGLGKL	61
QY	63	VSTRQDIFKSSE--NLDNIVRSIPG-AFTQODK--SSGIVSLNIRGDSGFGFRVNTMVDGIT	118
Db	62	VKTADTSLKEQVLDIIRDLTRYDGIIVAVEQGRGASSG---YSIRG-MDKNRVSLTVDGLA	117
QY	119	Q-TFYSTSTDAQ--RAGGSSQFGASVDSNFIAGLDVYVYKSGFSAGINSLAGSANLRTLG	175
Db	118	QIQSYTAQAALGCTRTAGSSGAINETIYENVKAVEISKSGNSVEQSGGALAGSAVAFQTKT	177
QY	176	VDDVVOGNTTYGILLKGLTCTNKGMAAIGARKWLESGASGVLYGHRSRVAQNYR	235
Db	178	ADVVIGEGRQWGIQSK-----TAYSGKNRGLTQISIA	208
QY	236	VGGGGHIGNFGAEYLERRKQRYFVQEGGLKFNSNGKWERDFQR--PYWKTWYQKYND	293
Db	209	LAG--RIG--GAELLIIHTRH--AGEIRAHEDAGRQVQSFNRLVPVEDTHPYANFIV	260
QY	294	PQELQ-----KYIEGHDK-----SWRENLAPOYDITPIDPSSLKOQS-----AGN	333
Db	261	EEECESGNYETCKSKPKKDVVGKDERQTVSTRDYTGPNRFLA--DPLSYESRSLWFRPGF	318
QY	334	LF--KLEYDCVFNKYTAQ---PRDLNTKIGSRKII---NRNYQFNGLSUNSLANLTA	385
Db	319	RFEKNRHYIGGILHEHTQQTEDTRDMTVPAFLTKAVFDENRKY-----	360
QY	386	AYNSGRQKYPKSGKFTGWGLKDFETYNNAKILDLNNTATFRLPRETELOTTLGFNFHN	445
Db	361	GPIRGNGKAGADHKY--GGLFTNGE--NGALVGAEGYGTGVFY--DETHKSRGLEXYVT	414
QY	446	EYKGNRFPPEELGLFFD--GPDQDNGL-----YSYLGREFKDGKL	482
Db	415	NADKDTWVDYARLSYDRQIGLDNHFQOQTHCSADGSKYCRPSADKPFYS---YKSDRVI	471
QY	483	LPQKSTIVQAPAGSOYNTFYFDDAALKDIYRLNYSTNTVGYRFGG--EYTGYY-----	533
Db	472	YGESHLLQAAFKSFDTAKI-----RHNLSVNLGDFRDFDNLRHQDYIYQHANRA	522
QY	534	-----GSDDEFKRAFGENSPYKKHCNQSCGIEYEPVLKK-----	567
Db	523	YSSKTPPTANPNGGKQDPYVWSIGRGNVVYTGQICRLGNNTYTDCTPRSINGSKSYAAGR	582
QY	568	-----YCKKRANHHSVS-----ISADFGDYFMPFASYSRTH-----	598
Db	583	DNVRLGRWADVAGALRYDYRSTHSDGVSSTGTHRTLSWAGIVLKPOTDMLDLYRTSTG	642
QY	599	-RMPNQEMYFSGISGVHTALKPERANTWQFG-----FNTYKGLLKQD	643
Db	643	FRLPSEAEMYGWRAGVQSKAVKIDPEKSFNKEAGIVFKGDFGNLEASWFNNAYRDLIVR-	701
QY	644	DTLGLKLVGYSRDN-----YIH-----NVYKG--W---WDLNGNIP-SW	678
Db	702	-----GYEAIQNGKETAKDPAAYLNQASARITGINVLGKIDWNGWD---KLIPEGW	750
QY	679	VSTGTGLAYTIQH--RNPKDKVHKHGFLELNYDY--GRFTNLSYAYOKSTQPTNFSAS	734
Db	751	YST--FAYNRVHVVDIKKRADRTDIQSHL-FDAIQPSRYVVGLEY-----	792
QY	735	ESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTGRWLNKLTGLGAMRYFGKISRATAE	794
Db	793	-----DQPEGKGVNGMLT---YSKAKEITELLSRALLG-----	825
QY	795	ERYIDGTNGNTSNVQLGKRSTKQETLAROPLIIDFYAAVEPKKNLIFRAEVKNLFD	854
Db	826	-----NSRNTKATARTRPWYIVDVSIGYTIKKHFTLRAGVYNLLNY	867
QY	855	RYI	857
Db	868	RYV	870



Db 350 SDL-TFDINTSTFNI-KTTFVHELLFGLQWLKNT--RNT-----LMYDKSKVRKADYNY 399  
Qy 473 LGREFKDGKLLPOKSTIVQAGSQYFNTFYFDAALK-KDIYRLNYSTNTVGYRFGGEYTG 531  
Db 400 -GYFQ-----PYMPSGRQYTAFLQDQIKWKNI--FST---GVRY----- 436  
Qy 532 YGSDDEKRAFGENSTYKHKHCQSGIYEPVLKKGKRRNNHS--VSIADFGDYFM 589  
Db 437 -----DHINNIGKNLAL--KYNDISAG-----HDYSQKNYNGWSYVLGLNYDVNHLS 483  
Qy 590 PFASYSRTHRPNIQEMYFSQIGDSGVHTA---LKPERANTWQFGFNTYKKGLKQDDTL 646  
Db 484 LFTNFKTRAPVIDEQYETQFQSSYPATSLNLEKEMINQTRVGGIITLNHLEQENDAF 543  
Qy 647 GLKLVGYRSRIDNYIH-----NVYKWWDLNGNI-PSWVSSTGLAYTIOHRNEKDKVHKH 700  
Db 544 QFRTTYFNRGKNEIFKTRGVNCGVNAADTNKVCPIIEN-----YRNLPGYVIQ- 594  
Qy 701 GFELELNYDYGRFFTNLSAYAKSTQPTNFSDASESNNASKEDQLKQGYCLSRVSALPR 760  
Db 595 GAELAYYQSTYLEGEITYSVKGRDT-----SPRN-----PWGKTSTWIAEIPP 640  
Qy 761 DYGRLEVGTWLGKLTGLGAMRYFGKSIRA-----TABERYIDGTNGGNTSNVRQLGK 814  
Db 641 RKATTALGFNVPKYYLTVGWRFAEFVRQDRSPLSGDPKASSWSLPASRGYSLHNL----- 695  
Qy 815 RSIKQOTETLARQPLIFDYAAYEPK--KNLIFRAEVKNLFDRRY 856  
Db 696 -----FLSWSPAKIKGMNVKITVDNLFNRAY 721

Search completed: July 24, 2002, 08:54:36  
Job time: 527 sec



Result No.	Query			ID	Description
	Score	Match	Length		
1	2514.5	51.3	913	1	YC17_HAEIN
2	278	5.7	915	1	TBPL_NEIGO
3	269.5	5.5	908	1	TB12_NEIMB
4	266	5.4	744	1	HXC1_HAEIN
5	257	5.2	911	1	TB11_NEIMB
6	252	5.1	999	1	HGPF_HAEIN
7	243	5.0	943	1	LBPA_NEIMB
8	242	4.9	912	1	TBPL_HAEIN
9	241.5	4.9	1013	1	HGPA_HAEIN
10	241	4.9	993	1	HGBC_HAEIN
11	232.5	4.7	999	1	HGPF_HAEIN
12	229	4.7	1063	1	HGFL_HAEIN
13	226.5	4.6	944	1	LBPA_NEIMA
14	218.5	4.5	723	1	Y262_HAEIN
15	211.5	4.3	725	1	HXC2_HAEIN
16	209.5	4.3	810	1	HPUB_NEIMC
17	209	4.3	1077	1	HGPA_HAEIN
18	205.5	4.2	1084	1	HGPF_HAEIN
19	204.5	4.2	810	1	HPUB_NEIMA
20	192	3.9	1066	1	HGFC_HAEIN
21	192	3.9	1067	1	HGBB_HAEIN
22	186.5	3.8	1046	1	HPUA_HAEIN
23	185	3.8	999	1	HGPA_HAEIN
24	181.5	3.7	972	1	HGBB_HAEIU
25	180	3.7	614	1	FTUB_ECOLI
26	180	3.7	1567	1	ICSN_XANCT
27	179	3.7	676	1	HMR_YERPE
28	175	3.6	687	1	HMR_YEREN
29	173.5	3.5	972	1	HGPA_HAEIU
30	170.5	3.5	700	1	YNCD_ECOLI
31	169.5	3.5	729	1	FHUE_ECOLI
32	168.5	3.4	774	1	FECA_ECOLI
33	167.5	3.4	1258	1	ICSN_ERWHE

RESULT	1	
YC17_HAEIN	STANDARD;	PRT; 913 AA.
ID	YC17_HAEIN	
AC	P45114;	
DT	01-NOV-1995 (Rel. 32, Created)	
DT	01-NOV-1995 (Rel. 32, Last sequence update)	
DT	16-OCT-2001 (Rel. 40, Last annotation update)	
DE	Probable tonB-dependent receptor HI1217 precursor.	
GN	HI1217.	
OS	Haemophilus influenzae.	
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;	
OC	Haemophilus.	
OX	NCBI_TaxID=727;	
ON	[1]	
RP	SEQUENCE FROM N.A.	
RC	STRAIN=RD / KW20 / ATCC 51907;	
RC	MEDLINE=95350630; PubMed=7542800;	
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,	
RA	Kerlavage A.R., Bult C.J., Tomb J.-P., Dougherty B.A., Merrick J.M.,	
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,	
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,	
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,	
RA	Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Botton R.C.,	
RA	Fine L.D., Fritchman J.L., Fuhrmann J.B., Geoghagen N.S.M.,	
RA	Gnehm C.C., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,	
RA	Venter J.C.;	
RT	"Whole-genome random sequencing and assembly of Haemophilus	
RT	influenzae Rd.";	
RL	Science. 269:496-512(1995).	
RN	[2]	
RP	IDENTIFICATION BY MASS SPECTROMETRY.	
RX	MEDLINE=20137488; PubMed=10675033;	
RA	Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,	
RA	Gray C., Fountoulakis M.;	
RT	"Two-dimensional map of the proteome of Haemophilus influenzae.";	
RA	Electrophoresis 21:411-429(2000).	
CC	-1- FUNCTION: PROBABLE RECEPTOR, TONB-DEPENDENT.	
CC	-1- SUBCELLULAR LOCATION: Outer membrane (Potential).	
CC	-1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.	
CC	-----	
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration	
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -	
CC	the European Bioinformatics Institute. There are no restrictions on its	
CC	use by non-profit institutions as long as its content is in no way	
CC	modified and this statement is not removed. Usage by and for commercial	
CC	entities requires a license agreement (See <a href="http://www.isb-sib.ch/announce/">http://www.isb-sib.ch/announce/</a>	
CC	or send an email to <a href="mailto:license@isb-sib.ch">license@isb-sib.ch</a> ).	
CC	-----	
CC	EMBL; U32801; AAC22870.1; --	
DR	TIGR; HI1217; --	
DR	InterPro: IPR000531; TonB_boxC	
DR	PROSITE; PS00430; TONB_DEPENDENT_REC.1; FALSE_NEG.	
DR	PROSITE; PS01156; TONB_DEPENDENT_REC.2; 1.	
KW	Outer membrane; Receptor; Signal; TonB box; Complete proteome.	
FT	SIGNAL 1 27 POTENTIAL.	
FT	CHAIN 28 913 PROBABLE TONB-DEPENDENT RECEPTOR HI1217.	

```
FT SITE 896 913 TONB C-TERMINAL BOX.
SQ SEQUENCE 913 AA; 102768 MW; 01BEFDE6D6AEF617 CRC64;

Query Match 51.38; Score 2514.5; DB 1; Length 913;
Best Local Similarity 51.38; Pred. No. 5.5e-151;
Matches 492; Conservative 155; Mismatches 229; Indels 83; Gaps 15;

Qy 1 MRSPRLKPICFYLMGVMLYHHSYAEDAGRAGSAQI-QVLEDDVHVAKRVPKDKKVFPTD 59
Db 1 MKKAIKLNIITLGLINTI-----GWTITQAQAEETLQDIDVEKVISNDKKPFTE 50

Qy 60 ARAYSTRODIPKSSSENLDNIIVSRIPGFTQODKSSGIVSLNIRGDSGFRVNTWVDGITQ 119
Db 51 AKAKSTREWFKEQTIIDVIRSIPIGFTQODKSGVSVNIRGENGLGRVNTWVDGVTTQ 110

Qy 120 TFYSTSTDAGRAGSSQFGASVDSNFTAGLVVKGSGSAGINSLAGSANLRLTGLVDDV 179
Db 111 TFYSTALDSGGSSQFGAAIDFNFTAGVDVKNFSFSGSAGINSALAGSANFRTLGVNDV 170

Qy 180 VOGNNTYGLLGLTGTNSTKGNAMAAIGARKWLESAGSVGLYGHRSRVAQNVRYGGG 239
Db 171 ITDKPFGIILKMGTSNATKSNFTWMAAGRWLDNGYGVVYGYSGREVSQDIIRI-GG 229

Qy 240 GQHIGFGEYLERKQRYFYOEGGLKFNSSNGKWERDQRPYWKTKW--YQKYND----- 293
Db 230 GERLASLQDILAKEKEAYFRNAGVIL--NPEGQWTPDLSKKHWCNKPQYKNGDCSY 287

Qy 294 -----PELOKY-----IEGDKSWRENLAPOQYDITPIDPSS 325
Db 288 RIGSAAKTRREILOELTNGKKPKDIEKLQKNGDGIETDKSFERN-KQOYSVAPIEPGS 346

Qy 326 LKQASAGLFLKLEVDGKFNKTAQFRDLNTRKIGSRKINRNYQFNYSLSNLYANLTA 385
Db 347 LQSRSHLLKFEYGDHONLGAQLRDLNKGSRKTENRNYQVNFNNNSYLDLNLMA 406

Qy 386 AYNRGROKPKSGFTGWLKLDPEYNNNAKILDNNTATFRPRETELOTLGFNFYHN 445
Db 407 AHNIGTKTIYKGGFAGQWADKLITKNVANIVDINNHSHTFLPKEDLTKTLGFNFYTN 466

Qy 446 EYGNRPPEELGLFDDPDODNGLYSY--LGRFKGDKLLPQKSTIVOPAGSQYFNTYF 503
Db 467 EYSNRPPEELSLYNDASHDQGLYSHKGRYSGTSLLPQRSVILQPSGKQKFKTYVF 526

Qy 504 DAALKDILYRNYTNTVYGRFGEGYGYGSDDEKFRAGFENSPYTKKHCNQCSCGYEP 563
Db 527 DTALSKGYHLNYSVNTHFAYFGEYGY-----ENTAGQO-----INEP 566

Qy 564 VLKYGKRRANNHVSISADPDGYFMPFASYSRTHRPNTIOEMYFSOIGDSGVHTALKPE 623
Db 567 ILKSGHKAPNHSATLSAELSDFEMFPFTYSTRTHRPNTIQEMFQSQVSNAGVNTALKPE 626

Qy 624 RANTWQGFNTYKGLKQDDTLGLKLVGYRSRIDNYIHNVYKRWDLNIGNIFSWSSTG 683
Db 627 QSDTYQGFNTYKGLTQDDVLGVLGYRSFKTNYIHNVYGVWV--RDCMPTWAESNG 684

Qy 684 LAYTIQHRNFKDKVHKHGFLELNYDYGREFTNLISYAKSTOPTNFSDASESPPNASK 743
Db 685 FKYTIAHQNYKPIVKSGVELEINYDMGRFANVSAYQRTNPTNADASPNNASQ 744

Qy 744 DOLQOGGLSRVSLPRDYGRLEVTGRLWKNKLTGLGAMRYFKSIRATAERYIDTNG 803
Db 745 DILQOGGLSRVSLPRDYGRLEVTGRLWKNKLTGLGAMRYFKSIRATAERYIDTNG 803

Qy 804 GNTSNRQLGKRSIKQETLARQPLIFDEYAAVEPKKNIIFRAEVKNLFDPRRYIDPLDAG 863
Db 804 FKMTLRENYAYVKKTEDIKQPIILDLHVSYPEIKDLIIKAEVQMLDKRYVDPLDAG 863

Qy 864 NDAATQYSSDFDKDDEEVTNCNADKTLGNGKYGTGSKSVLTNFAFRGRTFLITMSYKF 922
Db 864 NDAASQRYSSSL-----NNSIECAQDSSAC-----GGSDKTVLYNFAFRGRTYILSNYKF 913
```

```
RESULT 2
TBPI_NEIGO
ID TBPI_NEIGO STANDARD; PRT; 915 AA.
AC Q01996;
DT 01-OCT-1993 (Rel. 27, Created)
DT 01-OCT-1993 (Rel. 27, Last sequence update)
DE 01-NOV-1995 (Rel. 32, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN Tbpl.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RC SEQUENCE FROM N.A.
RX MEDLINE=92394880; PubMed=1325963;
RA Cornelissen C.N., Biswas G.D., Tsai J., Paruchuri D.K.,
RA Thompson S.A., Sparling P.F.;
RT "Gonococcal transferrin-binding protein 1 is required for transferrin
RT utilization and is homologous to TonB-dependent outer membrane
RT receptors.";
RL J. Bacteriol. 174:5788-5797(1992).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M96731; AAA25503.1; -.
DR PIR; A43335; A43335.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC_1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 24
FT CHAIN 25 915 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 38 45 TONB_BOX.
FT SITE 898 915 TONB_C-TERMINAL_BOX.
FT SEQUENCE 915 AA; 102213 MW; 697CF74B1010422F CRC64;

Query Match 5.7%; Score 278; DB 1; Length 915;
Best Local Similarity 21.4%; Pred. No. 5.1e-10;
Matches 224; Conservative 127; Mismatches 323; Indels 374; Gaps 58;

Qy 5 FRUKPICFYLMGVMLYHHSYAEDAGRAGSBAQTOVLEDDVHVAK--RVPKDKKVFPTDARA 62
Db 7 FRLNILLCSLMTAL---PAYAENV-QAG-QAQEKQLDTIQVAKKQKTRDRNEVTGLGKL 61

Qy 63 VSTRQDIFKSS-NDLNIIVSRIPG-AFTQODK--SSGIVSLNIRGDSGFRVNTWVDGIT 118
Db 62 VKTADTLRSKEQVLDIIDLTRYDPGIAVVEOGRGASSG---YSIRG-MDKNRVSLTVDGLA 117

Qy 119 Q-TFYSTSTDAQ--RAGGSSQFGASVDSNFIAGIDVYVKGSGSAGINSLAGSANLRLTLG 175
Db 118 QIQSYTAQAALGCTRTAGSSGAINETIYENVKAVEISKGSNSVEQSGALAGSVAFTQTKT 177

Qy 176 VDDVQGNNTYGLLKL-GLTGTNS--TKGNAMAA-IGARKWL--ESGASVGLYGHRSR 229
Db 178 ADDVIGEGWGIQSGTAYSGKNRGLTQSTALAGRIGGAELLIRTRHAGEIRAH----- 233

Qy 230 VAQNVRYGGGQGHGFGAEYLERKQRYFYOEGGLKFNSSNGKWERDQRPYWKTKWYQ 269
Db 234 -----EAAGRGVOSFNRNLAPOVDGSGSYAFIVEECK-----NGGHEKCKRANP----- 276
```



```

QY 290 KYNDPQELQKYEIGHDK-----SWRENLAPOYDITPIDPSSLKQOS-----AGNLF--KLEY 339
Db 277 -----KDDVGEDKQRTVSTRTDYGPNRFLA--DPLSYESRSLWFRPFCFRFENKRHY 326
QY 340 DGVENKYTAQ---PRDLNKTIGSKRII-----NNYQFNYGLSLNSYANLNL 383
Db 327 IGGILERTQQTDFDRMTVPALTKAVFDANQKQAGSLRGNKYAGNH-----KYGLG-- 379
QY 384 TAAYNSGRQYPKGSKFTGGLLKDFTYNNAKILDLNNTATFRLPRETELQTTGLGNFYF 443
Db 380 ---FTSGENNAVPAGAEY-GTGVFYD-----ETHTKSRYGLEYV 413
QY 444 HNEYGKRFPEELGLFFD--GPDODNGL-----YSYLGRFKGDK 480
Db 414 YTNADKTDWADYARLSYDRQIGLDNHFQOOTHCSADGSDKYCRPSADKPFSY----YKSDR 470
QY 481 GLLPQKSTIVQAGSOYFNTFYDAALKDILYRLNYSTNTVGY-RFGG-----EYTGYYG 534
Db 471 VIYGESHKLLQAAPKKSFDATAI-----RHNLNVN-LGYDRFGSNLRHQDY--YYQ 518
QY 535 SDDEFKAFENSPTYKHKHNSQSGIYEPVLKYGKKRA-----NNHVSYSISAD----- 583
Db 519 SAN---RAYSLKTPP-----QNNGKKTSPNGREKNPYWVSIGRGNVVT 559
QY 584 ---FGD-----YMPFA-----SYSRTH----- 598
Db 560 QICLFGNTYDCTPRSRNGSKSYAAVRDNRVLRGRWADVAGLRYDRSTHSDGCVSTG 619
QY 599 -----RMPNIQEMYFSQIGDSGVHTALKPERANTWQ 629
Db 620 THRTLSNAGIVLKPADWLDITYRTSTGFLRPSFAEMYGWSRSGDKIKAVKIDPEKSPNKE 679
QY 630 FG-----FNTYKGLLKODDLGLKLGYRSDIN-----YIHH----- 662
Db 680 AGIVFKGDGFGNLEASWFFNAYRDLIVR-----GYEAQIKDGKEQVKGPNPAYLNAQS 730
QY 663 -----NVYGR--W---WDLNGNIP-SWVSSTGLAYT-IQHRNFKDKVHKHGFLELYND 709
Db 731 ARITGINILKIDWGVWD---KLPEGWYST--FAYNRVRVRDIKKRADRTDIOGSHL-FD 784
QY 710 YGREFTNLVAYQKSTQPTNFSDAESPNNASKEDQLKOGYGLSRVSALPRDYGRLEVGT 769
Db 785 -----AIQPSRYV-----VSGYDQPEGKGVNGMLT-----YSKAKEIT 819
QY 770 RWLGNKLTIGGAMRYFKSKSTRATAEERYIDTNGNITSNVRLQKRSIKOTETILARPLI 829
Db 820 ELLGSRALLNG-----NSRNTKATARTREPWYI 847
QY 830 DFYAAAYEPKKNLIFRAEVKNLFDRIYI 857
Db 848 VDVSGYTVVKKHFTLRAGVYNLLNHRIV 875

RESULT 3
TB12_NEIMB
ID TB12_NEIMB STANDARD; PRT; 908 AA.
AC Q06987;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPL.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 25-42.
RC STRAIN=CCUG 37603 / B16B6 / SEROGROUP B / SEROTYPE 2A;
RX MEDLINE=93345825; PubMed=8344530;
RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,
RA Jacobs E., Schryvers A.B.;
RT "Cloning and characterization of Neisseria meningitidis genes
RT encoding the transferrin-binding proteins Tbp1 and Tbp2.";
```

```

RL Gene 130:73-80(1993).
RN [2]
RP SEQUENCE OF 25-45.
RC STRAIN=CCUG 37603 / B16B6 / SEROGROUP B / SEROTYPE 2A;
RX MEDLINE=93307625; PubMed=8319886;
RA Griffiths E., Stevenson P., Byfield P., Ala'Aldeen D.A.,
RA Barriello S.P., Holland J., Parsons T., Williams P.;
RT "Antigenic relationships of transferrin-binding proteins from
RT Neisseria meningitidis, N. gonorrhoeae and Haemophilus influenzae:
RT cross-reactivity of antibodies to NH2-terminal peptides.";
RL FEMS Microbiol. Lett. 109:85-91(1993).
CC -1- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION.
CC -1- SUBCELLULAR LOCATION: Outer membrane.
CC -1- INDUCTION: BY IRON STARVATION.
CC -1- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL; Z15129; CAA78831.1; -.
DR PIR; JN0819; JN0819.
DR PIR; S33154; S33154.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
DR Outer membrane; Receptor; Signal; TonB box.
KW SIGNAL 1 24
FT CHAIN 25 908 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 25 45 TONB BOX.
FT SITE 891 908 TONB C-TERMINAL BOX.
FT SEQUENCE 908 AA; 101583 MW; FE2FF4974CAC3C31 CRC64;
QY 5 FRLKPICFYLMGVMLYHHSTAEADAGRAGSEAOIOVLDDVHVAK--RVPDKKVFETARA 62
Db 7 FRLNLCISLMTAL---PVYAEVY--QAEQAQEKQLDTIQVAKKQKTRRDNEVYGLKL 61
QY 63 VSTRQDIFKSSE-NLDNIVRSIPG-AFTQDDK--SSGIVSLNIRGDSGFGRVNTWVGIT 118
Db 62 VKSSDTLSKQOVLNIRDLTRYPGIAVVEQGRGASSG---YSIRG-MDKNVRVSLTVDGVS 117
QY 119 Q-TEYSTSTDAQ--RAGGSSQFGASVDNSFTAGLDVVVKVSGSGAGINSLAGSANLRTLG 175
Db 118 QIOSYTAQAALGGTTRTAGSSGAINIEYEVNKAVEISKGSNSSEYNGALAGSAVFQTKT 177
QY 176 VDDVVGNNYVGLLLK-GLGTNTSTKGNAMAAICARKWLESGASGVLYGHRSRVAQNY 234
Db 178 AADIIGEGKQWGIQSKTAYSGKDHALTQSLALAGR---SGGAELLITTKRRGREIHAAH 233
QY 235 RVGGGGQHIGNFGAEYLERKKQ-----RYFVQE---GGLKFNSN-----SGKWERDF 278
Db 234 KDAGK--VQSFNRLVDEDEKKEGSGYRYFIVEEECHNGYAAACKNKLKEDASVKYDER-- 289
QY 279 QRPYWKTKWYQKYNDPQE-LQYIEGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKL 337
Db 290 -----KTVSTQDYTGSRNRLNLANPLEYGSQWL--FRGW-----HLDNR 326
QY 338 EYDGVFNKYTAQFRDLNLTIGSRKRIINRNYQFNYGLSLNSYANLNLTAAYNSGRQYPK- 396
Db 327 HYVGAVLERTQQTEDTRDMTVPAYFTSEDYVPGSLKGLGKY-----SGDKNAERL 376
QY 397 -----GSKFTGWGLLKDFEYNNNAKILDLNNTATFRLPRETELQTTGLGNFYHNEYGK 451
```

Query Match 22.5% Score 269.5; DB 1; Length 908;  
 Best Local Similarity 21.0%; Pred. No. 1.7e-09;  
 Matches 216; Conservative 116; Mismatches 354; Indels 343; Gaps 49;

```
Db 377 FVQEGSTLOGIG-----YGTGVFYD-----ERHTKNRYGVVEVYHNAKDT 418
QY 452 FPEELGLEFD--GPDODNGL-----YSYLGRFKGDKGLLPQKST 488
Db 419 WADYARLSYDROGIDLDNRLOQTHCSHDGSKNCRPGDKPYSF---YKSDRMIIYESRN 475
QY 489 IVQAGSGYFNT-----FYFDAAL-----507
Db 476 LFOAFVKAFDIAKIRHNLINLGYDREKSQLSHSDYQLQNAVQAYDLITPKKPPFNGS 535
QY 508 KKDYLRLNYSNTWG---YRFGGE-YT-----GYGSDDEPKRAFGENSPYTKK 552
Db 536 KDNRYRVSIGTKTNTVSPICRFGNNTYDCTPRNIGNGYY-----AAVQDNVRLGR 587
QY 553 HCNOSCGI-YEPVLKKYCKKRNHNSVS-----ISADFGDYFMPFASYSRTH-----R 599
Db 588 WADVGAGIRYD-----YRSTHSEDKSVSTGTHRNLNWSNAGVVLKPFMTWMDLTYRSTGFR 642
QY 600 MPNIQEMYFSQIGSGVHTALKPERANTWQFG-----FNTYKKGLLKQDDT 645
Db 643 LPSAEMYGWAGESLTKLDLKPESFNREAGIVFKGDFGNLEASYFNNAVYRDLI-----697
QY 646 LGLKLVYRSIDN-----YIH-----NVYK--WMDLNGNIPWSVSTGL 684
Db 698 ----AFGYETRTQGTQSASGDPGRYNAQNAARIAGINILGKIDWHGVWGLPDGLYST-L 752
QY 685 AY-----TIQHREKDKVHKHGFELNVDY--GREFTNLSYAYOKSTQPT 728
Db 753 AYNRKVKDADIRADRTFVTSYLFDAQVPSRYVLGLGYDHPDGWINGINTWMTYSK-----807
QY 729 NFSDASESPNNASKEDQLKQGYGLSRVSALPRDYGRLEVGTRWLGKLTLLGGAMRYFGKS 788
Db 808 -----AKSVDEL-----LGSQALLNG-----823
QY 789 IRATAERYIDGTNGGNTSVNRQLGKRSIKOTETLARQPLIFDYAYAEPKKNLIFRAEV 848
Db 824 -----NANAKKAASRR-----TRPWYVTDVSGYNTKTKHLLTRAGV 859
QY 849 KNLFDRRYI 857
Db 860 YNLLNRYV 868

RESULT 4
HCX1_HAEIN STANDARD; PRT; 744 AA.
AC P44523;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
GN Heme/hemopexin utilization protein C precursor.
OS HKUC OR H10113.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton R., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
```

```
CC -!- FUNCTION: REQUIRED FOR UTILIZATION OF FREE HEME AT LOW
CC CONCENTRATIONS (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -!- SIMILARITY: LOW, TO THE CORRESPONDING PROTEIN IN SEROTYPE B OF
CC H.INFLUENZAE.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32696; AAC21789.1; -.
CC TIGR: H10113; -.
CC InterPro: IPR000531; TonB_boxC.
CC Pfam: PF00593; TonB_boxC; 1.
CC PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Transport; TonB box; Signal; Receptor;
CC Complete proteome.
CC SIGNAL 1 24 POTENTIAL.
CC FT CHAIN 25 744 HEME/HEMOPEXIN UTILIZATION PROTEIN C.
CC SQ SEQUENCE 744 AA; 85043 MW; 74D94F72E41AEC31 CRC64;
CC -----
Query Match 5.4%; Score 266; DB 1; Length 744;
Best Local Similarity 20.5%; Pred. No. 2.2e-09;
Matches 181; Conservative 132; Mismatches 353; Indels 218; Gaps 39;
QY 14 LMGMVLYHHSYAEADAGRAGSEAOIQVLEVDVHVRAKRPDKKVFETDARAVSTRODIFK-S 72
Db 15 LVGIQL-NVTAKQSSNSNREELLPIIVTNDSDNKLKLP-----GRSVLKQKNIQXQ 65
QY 73 SENLDNIVRISPGAFTOODKSSGIVSLNIRGDSGFRVNTWVGITOTFTYSTSTADAGAG 132
Db 66 ADNAANLINILPGVNMAGFRPGGQTLNMGMDAEDVRVQLDCAKTSFEK-----116
QY 133 GSSQFGASVDSNFTAGLDVVKGFSFGSAGINSLAGSANLRTGLVDVGVQNTYVGLLKG 192
Db 117 -YQGSIFIEPELLRRVTVDKGNYSPOYGNGGFGAGVFTKDKARDPFLQENKIGGFLK- 174
QY 193 LTGTNSTKNGMAAIGARKWLESGASVGLYHSHRSRVAQNYRVGGGQHGIFGAEYLE 252
Db 175 -----YGNNSNNQKTYST-----ALVLQ 193
QY 253 RRQRYFVQEGGLKFNFS--NSGKWERDFQRPYWKTKWYQYNDPQELQY-----300
Db 194 NEQKNIDL-----LLFGSVRNAG-----DYKRPNDSKILFSKNNQTKGLIKLNQWISPEHLL 245
QY 301 ----IEGHDKSWRENLAPOYDITPIDPSSLKQOSAGNLFKLEVDGVFNKYTAQFRLDNTK 356
Db 246 TLSVYGIHKGW-EPPAAKRDLIP-----KPSLSIDIMRYGTD-IANKRLIYRD-----292
QY 357 IGSRKIIINRYQNFYGLSLNSYANLNTAAAYNSGRQK--YPKGSKFTGWGLL--KDFET 412
Db 293 ---QKDNVYTLKYNLPENNPNWLNSTQFSYKSTQNDMRPKKEASSGLVSLGNSQSWITY 349
QY 413 NNAKILDNLNTATFRLPRETELQTTLCFNTHNEFYHNEYGKRNFPPEELGLFFDQDNGLYSY 472
Db 350 SDL-TFDINNTSTFNI-KTTVHELLFGLWLKNT--RNT-----LMYDKSKVRKADYNY 399
QY 473 LGRFKGDKGLLPQKSTIVQAGSQYFNTFYFDAALK-KDIYRLNYSNTYGVYRFGGEVYG 531
Db 400 -GTFQ-----PYMPSGRQYTOAFYLDQOIKWNKII---FST---GVRY-----436
QY 532 YGSDDEKFRAGFGENSPYTKKHNCQSGIYEPVLKYYKKGKRNHNS--VSISADFGDYFM 589
Db 437 ----DHINNIGQKNLAL--KYNDISAG-----HDYSQKNYNGWSYLLGLNVDVNHLS 483
QY 590 PFASYSRTHRMPNIQEMYFSQIGSDSGVHTA---LKPERANTWQFGFNTYKKGLLKQDDTL 646
```

```
Db 484 LFTNFSWTRAPVDEQYETQFQSSVPAYSLSNLEKEMINQTRVGGIITLNHLFQENDAF 543
QY 647 GLKLVGRSRIIDNYIH-----NVYKGMWDLNGNI-PSWVSSTGLAYTIOHRNFKDKVHKH 700
Db 544 QFRTTYFNRGKNEIFKTRGVNCVGNAAADTNKVCPIIEN-----YRNLPGVYIQ- 594
QY 701 GFELELYNDYGRFTNLISYAYQSTQTFNFSADSESNASKEDOLKQGLSRVSLPR 760
Db 595 GAELIAYQSTYLFGETYSYVGRKDT-----SPRN-----PMWKTSTWIAEIPP 640
QY 761 DYGRLEVGTRWLNKGLTGLGAMRYEGKSIRA-----TAERYIDNGNGNTSNVRQLGK 814
Db 641 RKATTALGFNVPKYLIVGNRAFEVRQDRSPKASSWSLPASRGISLHNL----- 695
QY 815 RSIKQETLARQPLIFDYAYEPEK---KNLIFRAEVKNLFDRRY 856
Db 696 -----FLSWSPAKIKGMNVKITVDNLFNRAY 721

RESULT 5
TBII_NEIMB
ID TBII_NEIMB STANDARD; PRT; 911 AA.
AC Q09056;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Transferrin-binding protein 1 precursor.
GN TBPI.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A., AND PARTIAL SEQUENCE.
RC STRAIN=CCUG 37608 / M982 / SEROGROUP B / SEROTYPE 9;
RX MGLDLINE=93345825; PubMed=8344530;
RA Legrain M., Mazarin V., Irwin S.W., Bouchon B., Quentin-Millet M.-J.,
RA Jacobs E., Schryvers A.B.;
RT "Cloning and characterization of Neisseria meningitidis genes
RT encoding the transferrin-binding proteins Tbp1 and Tbp2.";
RL Gene 130:73-80(1993).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: BY IRON STARVATION.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z15130; CAA78833.1; -.
DR PIR: JN0821; JN0821.
DR PIR: S33156; S33156.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC.1.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; 1.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Receptor; Signal; TonB box.
FT SIGNAL 1 24
FT CHAIN 25 911 TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 38 45 TONB BOX.
FT SITE 894 911 TONB C-TERMINAL BOX.
FT SITE 911 911 MW: 99283ABAEUB773E6 CRC64;
SQ SEQUENCE 911 AA; 101631 MW; 99283ABAEUB773E6 CRC64;
```

Query Match 5.2%; Score 257; DB 1; Length 911;  
Best Local Similarity 21.5%; Pred. No. 11e-08;  
Matches 218; Conservative 131; Mismatches 356; Indels 308; Gaps 55;

```
RESULT 6
HGPB_HAEIN STANDARD; PRT; 999 AA.
ID HGPB_HAEIN
```

```
QY 5 FRKPICFYLMGVMLYHHSYAEDAGRAGSEAAQIOVLDEHVHAK--RVPKDKKYFTDARA 62
Db 7 FRLNILSLMTAL---PAYAENV-QAG-QAQEKOLDTIOVKAQKQKTRRDNVETGLKGL 61
QY 63 VSTRQDIFKSSSE-NLDNIVRSIPG-AFTQODK---SSGIIVSLNIRGDSGFGFVNMTVDGIT 118
Db 62 VKTADTUSKQOVLDIRLDTRIDPGIAVVEQGRASSG---YSIRG-MDKNRVSLUTVDGLA 117
QY 119 Q-TFYSTSTDAG--RAGGSSQFGASVDSNFIAGLDVYVKGFSFGSAGINSLAGSANLRTLG 175
Db 118 QIQSYTAQAALGGTRTAGSSGAINIEYENKAVEISKSGNSVEQSGSALAGSVAFOTKT 177
QY 176 VDDVQGNNTYGLLLKGLTGTNTSTKGNMAAIGARKWLESGASVGVIYGHHSRVSVAQNYR 235
Db 178 ADDVIGEGROWGIQSK-----TAYSGKNRGLTQSLA 208
QY 236 VGGGQGHIGNFGAEYLERRKQRYFVQEGGLKFNSNSCKWERDFOR--PYWKTKWYQYND 293
Db 209 LAG---RIG--GAELLIIHTGR---RAGEIRAHADAGRGVQSFNRLVPVEDSSEYAFIV 260
QY 294 PQELQ-----KYIEGHDK---SWRENLAPOYDITPIDPSLKKQOS-----AGN 333
Db 261 EDECEGKNYETCKSKPKKDVGVNDERQTVSTRDYTGPNRFLA--DPLSYESRSWLFPRGF 318
QY 334 LF--KLEYDGVFNKYTAQ---FRDLNTYKIGSRKIINRYQFNGLSLSNYSANLNLTAAYN 388
Db 319 RFENKRHYIGGILEHTQOTFTDRTMTVPAFLTKAV-----FDANSKQAGSLPGNGKY- 370
QY 389 SGROKYPKSGKFTGW--GLLKDFETYNNAKLTLNLNTATFRLPRETELOITLGFNYEHNE 446
Db 371 AGNHKY--GGLTNGENGALVGAE-YGTGVFYD-----ETHTKRGYLEYVYTN 416
QY 447 YGKNRPPEELGFDF--GPDQDNGL-----YSYLGRFKGDKGLL 483
Db 417 ADKDTWADYARLSVDROGIGLDNHFOQTHCSADGSKYCRPSADKPSY---YKSDRVY 473
QY 484 PQKSTIVQAPAGSQYFNTF-----YFDAALK-KDIYRLN-----YSTWT----- 520
Db 474 GESHRLQAQAFKSFDTAKIRHNLVNLGDFRSDNRHODYVYOHANRAYSSKTPPKTA 533
QY 521 -----VGYRFGGYTG---YYGSD---DEFKATGENSEPTYKKHCNQCQGIYEP 563
Db 534 NPNGDKSKPVWSIGGNGVVTGQICLFGNNYTDCTPRSNGKSYAAVRDNVRLGRWAD 593
QY 564 V----LKKYGKKRANHSVS-----ISADFGYFMFASYSRTH-----RMPNIQEMY 607
Db 594 VGAGLRVDYRSTHSDGCVSTGTHRTLSSWAGIVLKPADWLDLTYRTSTGFRLPSPAEMY 653
QY 608 FSQIGDSGVHTALKPERANTWQFG-----FNTYKGLLKQDDTLGLKLVGY 653
Db 654 GWRSGVQSKAVKIDPEKSPNKEAGIVFKGDFGNLEASWFNNAYRDLIVR-----GY 704
QY 654 RSRIDN-----YIH-----NVYKG--W---WDLNGNIP-SWVSSGTGLAYTI 688
Db 705 EAQIKNGKEAKGDPAYLNAQASARITGINIGKIDWNGVWD---KLPEGWST--FAYNR 759
QY 689 QH-RNEFKDKVHKHGFLELNYDY---GRFTNLISYAYQSTQPTNFSADSESPNNASKED 744
Db 760 VHVRIKKRADRTDIQSHL-FDAIQPSRYVVGLGY-----D 794
QY 745 OLKQGYGLSRVSLPRDYGRLEVTGRWLNKGLTGLGAMRYFGKSIRATAERYIDGTNGG 804
Db 795 QPEQKGVGVNGMLT-----YSRAKEITELLSRALING----- 826
QY 805 NTSNVRQLGKRSIKQETLARQPLIFDYAYEPEKKNLIFRAEVKNLFDRRYI 857
Db 827 -----NSRNTKATARRTPWYIVDVSGYYTIKKHFTLRAGVYNLLNRYV 871
```



```

Db 772 HNDWGFSTLFTKYNKFNIDLIIFGKND----FPLVSGGSLPFLSLYNINRDNASLK 827
QY 701 GFELELNLYDGR-----FFTNLSYAYQKSTOPTNFSDASESPNNASKEDQLKQCYGLS 753
Db 828 GIEINSKVLGKMAKFMGDFGNLSKYTYQKGRMGNI-----PMNAIQPRTWVYGLGYD 881
QY 754 RVSALPRDVRLEVGTRWLGKNTLGLGAMRYFGKSTRATAERYIDGTNGGNTSNVRQLG 813
Db 882 HPN-----HKGFDFTT-----HVKNPEDTIDYAKD 911
QY 814 KRISIKOTET----LARQPLIFDYAAYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQ 869
Db 912 K---KOTDSIKWRSKSYTILDIGVQPKNTLRAGYNLTNRKYIT-----WDSA-- 961
QY 870 RYSSFDPKDKDEVTCNADKTLGCKYGGTSGVLTNFARGRTFLITMSYKF 922
Db 962 RSIRSGTNSVIEQ-----STGLGINRFYAPGRNYSKMSVQFEF 999

RESULT 7
LBPA_NEIMB
ID LBPA_NEIMB STANDARD; PRT; 943 AA.
AC Q06379; Q9JYK5;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Lactoferrin binding protein A precursor (Iron-regulated outer membrane protein A).
GN LBPA OR IRGA OR NMB1540.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=BNVC / SEROGROUP B;
RA MEDLINE=94011384; PubMed=8406871;
RX Petterson A., van der Ley P., Poolman J.T., Tommassen J.;
RT "Molecular characterization of the 98-kilodalton iron-regulated outer membrane protein of Neisseria meningitidis.";
RL Infect. Immun. 61:4724-4733(1993).
RN [2]
SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwinn M.L., Deboy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B., Cotton M.D., Utterback T.R., Khouri H., Qin H., Vamathevan J., Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L., Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;
RA "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.";
RT Science 287:1809-1815(2000).
RL Science 287:1809-1815(2000).
CC -!- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- INDUCTION: BY IRON STARVATION.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).
CC
CC
CC EMBL; X69214; CA49148.1;
CC EMBL; AE002504; AAF41895.1;
CC TIGR; NMB1540;
CC InterPro; IPR000531; TonB_boxC.

```

```

Pfam: PF00593; TonB_boxC: 1.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE; PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor; Complete proteome.
FT SIGNAL 1 27 POTENTIAL.
FT CHAIN 28 943 LACTOFERRIN BINDING PROTEIN A.
FT SITE 826 943 TONB C-TERMINAL BOX.
FT CONFLICT 8 8 P -> Q (IN REF. 1).
FT CONFLICT 16 18 IAT -> VAA (IN REF. 1).
FT CONFLICT 22 22 A -> S (IN REF. 1).
FT CONFLICT 26 31 QAGAT -> NPETAA (IN REF. 1).
FT CONFLICT 43 43 I -> V (IN REF. 1).
FT CONFLICT 64 64 V -> A (IN REF. 1).
FT CONFLICT 233 233 R -> H (IN REF. 1).
FT CONFLICT 243 243 E -> A (IN REF. 1).
FT CONFLICT 247 247 D -> N (IN REF. 1).
FT CONFLICT 257 269 DIKRTREFFSV -> GIKKPSGEGYFLA (IN REF. 1).
FT CONFLICT 273 275 RES -> SEL (IN REF. 1).
FT CONFLICT 281 281 L -> V (IN REF. 1).
FT CONFLICT 284 286 YGK -> NGN (IN REF. 1).
FT CONFLICT 313 313 Q -> M (IN REF. 1).
FT CONFLICT 389 389 E -> K (IN REF. 1).
FT CONFLICT 420 423 KNLV -> OKLI (IN REF. 1).
FT CONFLICT 455 455 A -> K (IN REF. 1).
FT CONFLICT 546 546 K -> N (IN REF. 1).
FT CONFLICT 564 572 STGFENNQ -> YSDYTDKG (IN REF. 1).
FT CONFLICT 658 658 L -> V (IN REF. 1).
FT CONFLICT 667 667 V -> L (IN REF. 1).
FT CONFLICT 943 AA; 105680 MW; BD569ECACFC01A84 CRC64;
SEQUENCE 943 AA; 105680 MW; BD569ECACFC01A84 CRC64;

Query Match 5.0%; Score 243; DB 1; Length 943;
Best Local Similarity 20.3%; Pred. No. 8.5e-08;
Matches 224; Conservative 150; Mismatches 375; Indels 352; Gaps 58;

QY 9 PICFYLMGVMLYHHSVAEDAGRAGSE-AIQVLVEDVHVAKRVPKKVETDRAVSTRQ 67
Db 8 PLTLTALATATAPPAYAAQAGGATPDAAQTQSLKEITVRAAKVGRRSKEATGL-----G 61
QY 68 DIPKSSENLD-----NIVRSIPGAFTQODKSSGIVSLNIRGDSGFGVRNMTVDGITQ 119
Db 62 KIVTSETLNKEQVLGIRDLTRYDPGVAVVVEQNGASGGYSIRGVOK-NKRVAVSDVQVAQ 120
QY 120 ----TFYSTSTADAGAGSSQFGASVDNFIAGLDVYVKGFSGSAGINSLAGSANLRTLG 175
Db 121 IQAFTVQGSLSGYGGRGSGAIN-EIEYENISVTEIDKGAGSDHSGSGALGGAVAFRTKE 179
QY 176 VDDVVOGNNTYGLLLKGLTGNTSTKGNMAAIGA-----RKWLES-----G 216
Db 180 AADLISDGKSWGIOAK--TAYGSKNRQFMKSLGAGFSKDGWEGLLIINTERQRETRPHGD 237
QY 217 ASVGVLVGHSSRRSVAQRYNRVGGGGQHGNFGAAYLERRKOR--YFVQEGGLKFNSNSGW 274
Db 238 IADGVEYIDR-----LDAPQTYDIKRTREFFSVEG----- 271
QY 275 ERDFQRPYMKTKWYQKYNPDQELQKIEGHDKSWRENLAAPQYDITPIDPSLKKQOQSAGNL 334
Db 272 ERE-SKPVAKLAGYGYKLYNNQ-LNRWVKER-----IEQNPLSAEED----- 311
QY 335 FKLEYDGVFNKYTAQPFDLNLTGKSRKINRNQYFNGLSLNSYAN----LNITAYNSG 390
Db 312 -----AQVREAQAR-----HENLSAQAYTGGGRLTDPDPMYRSG 345
QY 391 RQKYPKGSKFTG---WGLLKDFETYNNAKILDLNNTATFRLPRETELQTLGFNFYHNE 446
Db 346 SWLAKLGYRFGGRHYGVGVFED--TKQRDYDIRM-----TEKQ-----YIGTD 386
QY 447 YGKNRFPPEELGLFFDGDQDNGLY--SYLGRFRKGDKLLPQKSTIVOPAGSOYFNTFYFD 504
Db 387 EAE-KFRDKSGV-YDGDGDFRDLGYFVFNIEWKGNL-----VRGIGLKYSRTKFID 437

```

```
QY 505 AALKKDIYRLNVTNTVG---YRFGGEYTYGYSDDF-----KRAFGENSPYKKHQNQC 558
Db 438 E-----HHRRMGLLYRENAEYSDNWADKAVLSPDKQGVATDNNTLK-----LNC 484
QY 559 GIPEVLKYGKRRANHSSISADGDDPMPFASYSRTHRMFNQIEMFESQ-----610
Db 485 AVY-PAVDKSCRASADK---PYSYDSDRF---HYREQH---NVLNASFEKSLKNNKWK 533
QY 611 -----IGDSGVHTALKPER-----ANTWQFGNTYKKGILLKQDDTLGLKLGVY---653
Db 534 HHLTLGFGYDASKAISRPESQLSHNAARISESTGFDENNQDKYLLGKPEVGVSGVGIET 593
QY 654 -RGR--IDNVIH--NVY-----GKWD--LNGNIP-----SW 678
Db 594 LRSKVCPRKINGSNIHISLDRFSIGKVPDFSLGGRYDRKNFTTSEELVRSGRYVDRSW 653
QY 679 VSTGLA-----YTIQHR---NFKDKVHKHGFELNLDY-----710
Db 654 --NSGLFKNRHFVSYSRASSGFRTPSFQELFGIDYHDYDKGWORPALKSEKAANREI 711
QY 711 -----GRF-FTNLSYAYQKSTQPTNFSF-ASESPNASKEDQL-----KQYGLSRVS 756
Db 712 GLQWKDGFGLFESSFNRYTDMIAVDHKTLPNQAGQLTEIDIRDYNAQNNWSLQGVN 771
QY 757 ALPR-----DYGRLEVG--TRWLGKLTGGAMRYFGKSIR-----ATAEERYDG----800
Db 772 ILGKIDWNGVYGLPEGLYTTLAYNRKPKSVSNRPGLSLRSYALDAVQPSRYVLGFGYD 831
QY 801 -----TNGGNTSNVROL-GKRSIKQETLARQPLIFDFYAAEYEPKKNLIER 845
Db 832 QPGKKGANIMLYSGKGNPDELAYLAGQKRYSTKRASSWSWADVSAYLNLUKRLTLR 891
QY 846 AEYKNLFDRIYDPLDAGNDAATQRYSSFPDKPKDDEEVTCDNADKTLGNCYGGTGSKSVL 905
Db 892 AAIYNGIYRYV-----TWESLRQTAESTAN-----RHGSDS-----923
QY 906 TNFAR-----GRTELIWYSKYF 922
Db 924 -NYGRYAAPGRNFSALENKF 943

RESULT 8
TBPL_HAEIN STANDARD; PRT; 912 AA.
AC P44970;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1995 (Rel. 32, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable transferrin-binding protein 1 precursor.
GN TBPA OR TBPI OR HI0994.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald J.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
CC -!- FUNCTION: ACTS AS A TRANSFERRIN RECEPTOR AND IS REQUIRED FOR
CC TRANSFERRIN UTILIZATION (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (Potential).
```

```
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: U32780; AAC22656.1; -.
CC TIGR: HI0994;
CC InterPro: IPR000531; TonB_boxC.
CC Pfam: PF00593; TonB_boxC; 1.
CC PROSITE: PS00430; TONB_DEPENDENT_REC.1; 1.
CC PROSITE: PS01156; TONB_DEPENDENT_REC.2; 1.
CC Outer membrane; receptor; Signal; TonB box; Complete proteome.
KW SIGNAL 1 23 POTENTIAL.
FT CHAIN 24 912 PROBABLE TRANSFERRIN-BINDING PROTEIN 1.
FT SITE 50 57 TONB BOX.
FT SITE 895 912 TONB C-TERMINAL BOX.
SQ SEQUENCE 912 AA; 103233 MW; 294C08991A652CC2 CRC64;

Query Match 4.9%; Score 242; DB 1; Length 912;
Best Local Similarity 20.2%; Pred. No. 9.4e-08;
Matches 213; Conservative 136; Mismatches 419; Indels 288; Gaps 49;

QY 5 FRLKPICFVLMG--VMLYHHSYAEDAGRASGEAQIQVLED-----VHVAKRV--PKDK 55
Db 7 FRLSIISCLLISCYKVAETQSIKDTKEAISSEVDYDTOSTDESELETISVTAEKYRDRKONE 66
QY 56 VETDARAVSTRODIFKSSSE-NLDNIVRSPGAFTQDKSGSIVSLNTRGDSGFRVNTMV 114
Db 67 VTGLGKIITSEISIREQVNLNRLTRYDPGISVVGRCASSGYISRG-MDRNRVALLV 125
QY 115 DGIOTQ-FYSTSTDAGRAGGSSQFGA--SVDSNFIAGLDVVKGSFGSGAGINSLAGSANL 171
Db 126 DGLPQTSYVWQSPVARSYSGTGAINEYENKAVEISKGSSEYNGALAGSVTF 185
QY 172 RTLGVDVVGQNTTYGLLLKGLTGNTSKG--NAMAAGARKWLESGASGVLYGHSRRS 229
Db 186 QSKSAADILEGDKSWGIGQTKN-AYSSKNKGFTHS LAVAG-----KGGFEGELAIYTORNS 239
QY 230 VAQNYRVGG--GGQHIGNFGAEYLERRKQRYVQE-----262
Db 240 IETQVIKDKALGVQSYNRLIAK--PENQSAFYVMEDEC PKGYDECIPSAKPAILSTKKE 297
QY 263 -----GGLKFNSSNGKWERDFQRPYKTKWKYQYNDPQELQKYIEGHDKSWREN LAP 314
Db 298 TVSVSDYTGANRIKPNPMKYE-----SQSWFLRGYHFSQHYIGGIFETQOQ---345
QY 315 QYDITPID-PSSLKQOSAGNLF-----KLEY-----DGVFNKYTA--QPRDLNTK--356
Db 346 KFDIRDMTFPAYLSPTKPGDLANRPYPKODYGAYQHIEDRGVGYVAGSLYDFDEHRRKOR 405
QY 357 -----IGSRKIINRYQFNYSGLNSYANLNLTAAAYNSGRQKYPKGSKFTG 402
Db 406 VGIEYIYENKNKAGIIDKAVLSANQQ---NIILDSYMR-----HTHCSLYPNPSKNCR 455
QY 403 WGLLKDFEYNNNAKILDLNNTATFRLPRETELQT-----TLGFNRYF-----HNEYG 448
Db 456 PTLDRPYSYRSDRNVYKEKHNMLQLNLEKKIQNNWLTHQIVFNGLGDFDDFTSALQHKDYL 515
QY 449 KNRFPPEELGLFED--GPDOONGL--YSYLGKPGKDGKLLPQKSTIVQPSAGSYFNTFYFD 504
Db 516 TRRVATADSISDKTGKTRNGLREYPYL-----YPKPKPYF-----552
QY 505 AALKKDIYRLNVTNTVGVYFGGEYTYGYSDDF-----FGNSPTYKK 552
Db 553 -----AGEDHCNYOGSSSNYRDCVKRLIKGNKYFAARNMALGK 592
QY 553 HCNQSGIYEPVLKYGKRRANHSSYSI-----SADFGDYMPF-----ASY--SRTHRM 600
```



QY 567 KYGKKRANNHVSISADGDFYFMPFASYSRTHRMPIQEMYS-QIGDSGV--HTALKPE 623  
Db 716 KNTWQFGNTYKGLLKQDDTLGLKLVGRSRIDNIHNVYKWKWDLNIGNISWSSSTG 683  
QY 724 RYAKTEKTAFT-----LHDDWGFISTSLFKTNRYDFIDLVIKGEREFVGNPN--NRGK 824  
Db 684 LAY-TIOHRNFKDKVHKHGELELNLYDGR-----FFTNLSYAYOKSQPTNFSASE 735  
QY 825 ISFTTFQIN--RDSAVVKGIEINSKVPFGMAKFMGDFNLISYKTYOKGRWDGNI----- 878  
Db 736 SPNNASREDOLKQGYLSRVSALPRDYGRLEVGTRWLGKNTLTLGAMRYFGKSTRATAEE 795  
QY 879 -PMNAIQPMVYGLGYDHP-----QKGFENFYTT--- 908  
QY 796 RYIDGTNGNTSNVRQLGKRSIKQET-----LARQPLIFDYAYEPKKNLIFRAEVKNL 851  
Db 909 -HVASKNPEDYDIYAKDK---NOTNTSIRKWSKSYTILDILIGYVQPIKNTIRAGVYNL 964  
QY 852 FDRYIDPLDAGNDAATORYSSDPKDKDEEYTCNADKTLNCKYGGTSKSVLTNEARG 911  
Db 965 TNRKYL---TWSARSIRSGFSNVDQK-----TGGINRFRYAPG 1002  
QY 912 RTFLITMSYKF 922  
Db 1003 RNYKMSVQFEF 1013

RESULT 10  
HGBC\_HAEIN STANDARD; PRT; 993 AA.  
AC Q9KIVO;  
DT 16-OCT-2001 (Rel. 40, Created)  
DT 16-OCT-2001 (Rel. 40, Last sequence update)  
DE 16-OCT-2001 (Rel. 40, Last annotation update)  
DE Hemoglobin binding protein C precursor.  
GN HGBC.  
OS Haemophilus influenzae.  
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;  
OC Haemophilus  
OX NCBI\_TaxID=727;  
RN [1]  
RP SEQUENCE FROM N.A.  
RC STRAIN=NTHI N182;  
RC MEDLINE=20316037; PubMed=10858226;  
RA Cope L.D., Hekal Z., Hansen E.J.;  
RT "Detection of phase variation in expression of proteins involved in  
RT hemoglobin and hemoglobin-haptoglobin binding by nontypeable  
RT Haemophilus influenzae";  
RL Infect. Immun. 68:4092-4101(2000).  
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OF THE HUMAN HOST AND  
CC IS REQUIRED FOR HEME UPTAKE.  
CC -!- SUBCELLULAR LOCATION: Outer membrane.  
CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE  
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA  
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAINING.  
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING  
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE  
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A  
CC WAY TO AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST.  
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;  
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.  
CC -----  
CC This SWISS-PROT entry is copyright. It is produced through a collaboration  
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -  
CC the European Bioinformatics Institute. There are no restrictions on its  
CC use by non-profit institutions as long as its content is in no way  
CC modified and this statement is not removed. Usage by and for commercial  
CC entities requires a license agreement (See [http://www.isb-sib.ch/](http://www.isb-sib.ch/announce/)  
CC or send an email to [license@isb-sib.ch](mailto:license@isb-sib.ch)).  
CC -----  
CC EMBL; AF221060; AAF80178.1; -

DR InterPro; IPR000531; TonB\_boxC.  
DR Pfam; PF00593; TonB\_boxC.1.  
DR PROSITE; PS00430; TONB\_DEPENDENT\_REC\_1; FALSE\_NEG.  
DR PROSITE; PS01156; TONB\_DEPENDENT\_REC\_2; 1.  
KW Outer membrane; Transport; TonB box; Multigene family; Signal;  
KW Receptor; Repeat.  
FT SIGNAL 1 24 POTENTIAL.  
FT CHAIN 25 993 HEMOGLOBIN BINDING PROTEIN C.  
FT DOMAIN 26 37 3 X 4 AA TANDEM REPEATS OF Q-P-T-N.  
FT REPEAT 26 29 1.  
FT REPEAT 30 33 2.  
FT REPEAT 34 37 3.  
FT SITE 47 54 TONB BOX.  
FT SITE 976 993 TONB C-TERMINAL BOX.  
SQ SEQUENCE 993 AA; 113616 MW; A551BF3B2C641612 CRC64;

Query Match 4.9%; Score 241; DB 1; Length 993;  
Best Local Similarity 19.8%; Pred. No. 1.2e-07;  
Matches 222; Conservative 156; Mismatches 414; Indels 328; Gaps 51;

QY 3 SSRLRPICILMGVMLYHHSYAEDAGRAGSEAQI-----QVLEDVHVK-----AK 48  
Db 2 ANFLANLAYSVMLGLTAGVAYAAQPTNQPTNQNSNVSEQLQINYSGSTENSDTK 61  
QY 49 RVPKDKVFTDARAVSTRODIFKSSSENLDNIVRSIPGAFQQDKSSGIVSLNIRGDSGF- 107  
Db 62 TPPIAETVKTAKTLEREQ-----ANNKDIVKETGV-----TJVAGRFGQSGFA 108  
QY 108 -----GRVNTMVDGITQTFYSTDAGRAG-----GSSQFGASVDSNFIAGLDVVK 153  
Db 109 IRGVDENRVAINIDGLRQA--ETLSQGFKELEFEGYFNFNTRNGAEIET--LKEVNITK 164  
QY 154 GFSGSGAGINSLAGSANLRTLGLVDDVVOGNNTYGLLLKGLTGNTSTKGNMAAIGARKWL 213  
Db 165 GANSIKSGSGSLGSGVYIKTKDARDYLLNKDYVYSYKGYATENNOSFNTLTLAGRYKKF 224  
QY 214 ESGASGVLYGHSRRSVAQNYRVGGGGQHIGNF---GAEYLERRRKQ---YFVQGG- 264  
Db 225 DA-----LVVTRRN-----GHELENDYKNADSLTOGKKREKADPYKIEQDST 268  
QY 265 -LKFNSNGKWER-----DF-----QRPYKTKWYQKYNDPOELQKYIEGHDKSWREN 311  
Db 269 LLKLSFNPTENHRFTLAADLYEHRSGQDLSYTLKYORSNETPEVSR--HTNDKTKRRN 327  
QY 312 LAPQYDITPIDPSSLKQOSAGNLFKLEYD-----GVF----- 343  
Db 328 ISFSYE-----NFSQTPFWDTLKLYSDQRIKTRARTDEYCDAGVTHCQGTENPTGLK 380  
QY 344 -----NKYTAQFRDLNLTIGSRK-----IINRNYQ-----FNYGLSLSYANLNL 383  
Db 381 LTNGKITRRNGTDLQFKEKCTAAGTKTYDFDTFIDTNQGEIVGKLNLYANDTWYDCSL 440  
QY 384 TAYNSGRQYKPKGSKFTGW-----GLLKDFETYN-----AKIL 418  
Db 441 FDCDKGTCKMKVFEGNATYGIKGRSELETEVLNCKRFARIKDRKDKNNRNRKIKSIL 500  
QY 419 DLNNTATERLPRTELQGT-----TLGFNYF---HN-EYKGN-----RPEELGLF 459  
Db 501 PSFEGYLERLWQRDLDTNTQQLNLDLTDFKTYRVEHNLQYGSSTYNTTKRNVNRAG-- 558  
QY 460 FDGPDQDNGLYSLGR-----FKG-----DKGLLPQ---KSTIVOPAGSY 497  
Db 559 YDATDVQWAKRTLKSHNFFTGQDIVETCATTTSSSAWNAFLCPRVDPPEFSLPIKTKE 618  
QY 498 FNTFYDAAALKDIYRLNLYSTNTVGYRFGG-EYTYGYG-----SDDEKRAF---G 544  
Db 619 KSYLFDNVVITD-----YLSFDLGYRYDNIHQPIKGVTPKLPDDIYKGLFPLPDK 673  
QY 545 ENSPTVKKHCNCGGIVEPVLKYYKKRANNSHVSISADFGYFMPFASYSRTHRMPIQ 604  
Db 674 PNSNQVKENVOQNLIDYIAKONKYY---KAHSYFASTIDTFSFLRLQLKYSKGFRAPTS 730



```

QY 605 EMYFS-QIGDSGV--HTALKPERANTWQGFNTYKGLLKQDD-----TLGLKLVGYSRI 657
DQ 731 EMYFTKHPDFTLLPNTNLKPEAKTEAKTFT-----LHNDWDGFISSLFKNTYRDFI 784
QY 658 DNYIHNVYKGMWDLNGLNIESWSSTGLATYIOHRNF-----KDKVHKGFELELNLDYGR- 712
DQ 785 D-LVH-----KTRFKVCKPSEGEIEFTDTONIRDSAVVKGIEINSKVLGKM 834
QY 713 -----FFNLSVAYQKSTQPTNFSNASKEDQLKQGYGLRSALPRDYGRLE 766
DQ 835 AKPMGDGNLSYKTYQKGRMDGNI-----PMNAIQPKTMVYGLGYDHPG----- 878
QY 767 VGRWLNGLITLGGAMRYGKSTRATAEERYIDGTGGNTSNVROLGKRSIKOTEN----- 822
DQ 879 -----OKFGFNFTT-----HVASKNPEDFYDIYAKDK-----NOTDTSIKW 915
QY 823 LARQPLIFDYAAYEPKKNLIFRAEYKNLFDRIIDYDLOAGNDAATQRYYSFDPKDKDE 882
DQ 916 RSKSYTILDLIGVQPKIKLITIRAGVYLNTRKYI-----TWDSARSIRSGTNSVIDQ-- 969
QY 883 EYTCNADKTLCLNGKYGKGTGSKSVLTNFARGTFLITWSYKF 922
DQ 970 -----STGGINRNFYAPGRNKMVSQVFEF 993

RESULT 11
HGP2_HAEIN
ID HGP2_HAEIN STANDARD; PRT; 999 AA.
AC P44809;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable hemoglobin and hemoglobin-haptoglobin binding protein 2
DE precursor.
DE GN HI0661.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA Mckenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
RN [2]
RP REVISIONS.
RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.,
RA Peterson J., Hickey E., Dodson R., Gwinn M.;
RA Submitted (MAY-1998) to the EMBL/GenBank/DBJ databases.
RN [3]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=20137488; PubMed=10675023;
RA Langer H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
RN [4]
RP CONCEPTUAL TRANSLATION.
RA Coudert E.;
RL Unpublished observations (AUG-2001).
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED

```

```

CC FOR HEME UPTAKE (BY SIMILARITY).
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAIRING.
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A WAY TO
CC AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR PROTEIN FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; TWO FRAMESHIFTS WERE
CC INTRODUCED IN THE REPEATS REGION AND IN THE LEADER PEPTIDE TO
CC MAXIMIZE THE SIMILARITY WITH OTHER ORTHOLOGS.
CC -----
CC THIS SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL; U32749; AAC22319.1; ALT_SEQ.
CC TIGR; HI0661;
CC InterPro; IPR000531; TonB_boxC.
CC Pfam; PF00593; TonB_boxC; 1.
CC PROSITE; PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE; PS01156; TONB_BOX; Multigene family; Signal;
CC Outer membrane; Transport; TonB box; Multigene family; Signal;
CC Receptor; Repeat; Complete proteome.
CC SIGNAL 1 24 POTENTIAL.
CC CHAIN 25 999 PROBABLE HEMOGLOBIN AND HEMOGLOBIN-
CC DOMAIN 26 53 7 X 4 AA TANDEM REPEATS OF Q-P-T-N.
CC REPEAT 26 29 1.
CC REPEAT 30 33 2.
CC REPEAT 34 37 3.
CC REPEAT 38 41 4.
CC REPEAT 42 45 5.
CC REPEAT 46 49 6.
CC REPEAT 50 53 7.
CC SITE 63 70 TONB BOX.
CC SITE 982 999 TONB C-TERMINAL BOX.
CC SEQUENCE 999 AA; 114690 MW; 1A17AAB220092B7D CRC64;
QY 40 LEDVHYK-----AKRVPKDKVFTDARAVSTRQDIFKSENLDNIVRSIPGAFTOODK 92
DQ 62 LEQINVSGSTENSQDTPPKPIAETVKTAKTLEREQ-----ANNIKDIVKYETGV----- 110
QY 93 SSGIVSLNIRGDSGF-----GRVNTWVDGITQTFSTSTADGRAG-----GSSQF 137
DQ 111 --TWVEAGRFQSGFAIRGVDENRVAINIDGLRQA--ETLSSQGFELFCGYGNFNTRN 166
QY 138 GASVDSNFIAGLDVVKVGSFSGSAGINSAGSANLRTGLVDVVGQNNYTGILLKGLGTN 197
DQ 167 GAETIET--LKEVNTIGADSIKNGSGSLGSGSVYKTKDARDYLINKDYYSYKGYATEN 224
QY 198 STKGNAMAAIGARKWLESASGVLYGH-----SRSVAQNYRGGGG- 240
DQ 225 NQSFDTLTLGRYKFKFDVLVVTTSRNGHELENYGKYNNDKIQGKKREKADPYKIEQDST 284
QY 241 -----QHIGNFGAEYLERKKQ-----RYPVQEGG-----LKFN 268
DQ 285 LKLSFNPTENHRFTFAADLYEHRSGQDLSYTLKY--QRSGNETPEVDSRHTNDKTKRR 342
QY 269 SNGSKWERDFQRYWKTKYQKYNDDQELQ-----KYIEGHDKSWRENLAPO 315

```

```

Query Match 4.7%; Score 232.5; DB 1; Length 999;
Best Local Similarity 19.8%; Pred. No. 4.2e-07;
Matches 209; Conservative 144; Mismatches 415; Indels 285; Gaps 47;
QY 40 LEDVHYK-----AKRVPKDKVFTDARAVSTRQDIFKSENLDNIVRSIPGAFTOODK 92
DQ 62 LEQINVSGSTENSQDTPPKPIAETVKTAKTLEREQ-----ANNIKDIVKYETGV----- 110
QY 93 SSGIVSLNIRGDSGF-----GRVNTWVDGITQTFSTSTADGRAG-----GSSQF 137
DQ 111 --TWVEAGRFQSGFAIRGVDENRVAINIDGLRQA--ETLSSQGFELFCGYGNFNTRN 166
QY 138 GASVDSNFIAGLDVVKVGSFSGSAGINSAGSANLRTGLVDVVGQNNYTGILLKGLGTN 197
DQ 167 GAETIET--LKEVNTIGADSIKNGSGSLGSGSVYKTKDARDYLINKDYYSYKGYATEN 224
QY 198 STKGNAMAAIGARKWLESASGVLYGH-----SRSVAQNYRGGGG- 240
DQ 225 NQSFDTLTLGRYKFKFDVLVVTTSRNGHELENYGKYNNDKIQGKKREKADPYKIEQDST 284
QY 241 -----QHIGNFGAEYLERKKQ-----RYPVQEGG-----LKFN 268
DQ 285 LKLSFNPTENHRFTFAADLYEHRSGQDLSYTLKY--QRSGNETPEVDSRHTNDKTKRR 342
QY 269 SNGSKWERDFQRYWKTKYQKYNDDQELQ-----KYIEGHDKSWRENLAPO 315

```

```
Db 343 NISFSYENFSPFPWDT-LKITYSD-QRIKTRARTEYCDAGVRHCEGTDNPTGLKVT-N 399
Qy 316 YDITPIDPSSLKQOAGNLFKLE---YDGVFNKYTAQFRDLNLTGKIGSKIIRNRYQENVG 372
Db 400 GKTRRDCGDLQFEKNTAKSSDKTYD--FKK-----FIDYDKRVIDDKLVNN----- 447
Qy 373 LSLSYANLNTAAYNSGRQYKPGSKETGNG-----LLKDFETVNNAKILDLNNTA 424
Db 448 -PSDTWYDCSIFCENNNAKIKVFGNNYGYDGKWKVEYDLEIKELNGKKFAKIKDNRKI 506
Qy 425 TFRLP-----RETELQ-----TLGFNFYHNE-----YG-----KNRPEE 455
Db 507 KSILPSSGYLERLWQERDLDTNTQQLNLDLTKDFKIHIEHNLQYGSYNTAMKVMNR 566
Qy 456 LGLFFDGPD-----QDNGLYSYLGR-----FKGDKGLLPQ-----KSTIVQAPAGSOFVN 499
Db 567 AG--NDASDVQWATPTLGEDSWTGKPHTCATTYEWANLCPRPVDPPEFSYLLPIKTKGS 624
Qy 500 TFYDAALKDIYRLNYSNTVGRFG-ETGYYG-----SDDEFKRAF-----GEN 546
Db 625 VYLFDFNVITD-----YLSFDLGYRYDNIHYOPRYKHGHTPKLPDDIVKGLFIPLPNSN 679
Qy 547 SPTYKKHCNQCSCGIEPVLLKRYGKKRANHSVISADEGDYFMPFASYSRTHRPNQEM 606
Db 680 SDPNKVENVOQNI-DYIAKONKYYKAHSYFVSTIDPTSELRQLKYSKGFRTPTSDM 738
Qy 607 YFS-OIGDSGV--HTALKPERANTWQGFNTYKKGKLLQDDTLGKLVGYSRSDINVIHN 663
Db 739 YFTFKHPDFTLPLNTDLKPEIAKTAFT-----LHNDWGFISTSLFKTYKNKFDL 792
Qy 664 VYGKWDLNGNIPSWVSTG---LATIQHNFKDKVHKHGFLELNYDGR-----712
Db 793 IFKKQETFK-----VGSGRGETLPESLYQNIINRDNASLKGIBINSKVLGKMAKPMFG 846
Qy 713 FFTNLSYAYOKSTOPTNFSDASESPNASKEDQLQCGYGLSRVSALPRDYGRLEVGTRWL 772
Db 847 FNLKYKTYQKGRMNGNI-----PMNAIQPRTVYGLGYDHPN-----884
Qy 773 GNKLTLGAMRYFKSIRATAERYIDGTNGGNTSNVRQLGKRSIKOTETL---AROPLI 829
Db 885 -----HKGFDFYTT-----HVASKNPEDTYNM--FYKEENKDKSTIKWRSKSYTI 928
Qy 830 FDFVAAVEPKNLFRAEVNLFDRYIDPLDAGNDAATORYYSSFPKDKDEVTNCAD 889
Db 929 LDLGYVQPKNLIRAGVNLNRKVI---TWDSARSIRSFCTSNVDO-----975
Qy 890 KTLGNGKYGTSKSVLNTNFARFGLTITMSYKF 922
Db 976 -----STGLGINRFYAPGRNKKMSVQFEF 999

RESULT 12
HGPI_HAEIN STANDARD; PRT; 1063 AA.
AC P44795;
DT 01-NOV-1995 (Rel. 32, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Probable hemoglobin and hemoglobin-haptoglobin binding protein 1
DE precursor.
GN HI0635.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
```

```
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus
RT influenzae Rd.";
RL Science 269:496-512(1995).
RN [2]
RP IDENTIFICATION BY MASS SPECTROMETRY.
RX MEDLINE=2013748; PubMed=10675023;
RA Langen H., Takacs B., Evers S., Berndt P., Lahm H.W., Wipf B.,
RA Gray C., Fountoulakis M.;
RT "Two-dimensional map of the proteome of Haemophilus influenzae.";
RL Electrophoresis 21:411-429(2000).
RN [3]
RP CONCEPTUAL TRANSLATION.
RA Couderc E.;
RL Unpublished observations (AUG-2001).
CC -!- FUNCTION: ACTS AS A RECEPTOR FOR HEMOGLOBIN OR THE
CC HEMOGLOBIN/HAPTOGLOBIN COMPLEX OF THE HUMAN HOST AND IS REQUIRED
CC FOR HEME UPTAKE (BY SIMILARITY)..
CC -!- SUBCELLULAR LOCATION: Outer membrane (By similarity).
CC -!- MISCELLANEOUS: THIS PROTEIN IS SUBJECT TO PHASE-VARIABLE
CC EXPRESSION ASSOCIATED WITH ALTERATION IN THE LENGTH OF THE CCAA
CC REPEAT REGION. THIS MECHANISM IS CALLED SLIPPED-STRAND MISPAGING.
CC ADDITION OR LOSS OF CCAA REPEAT UNITS WOULD CHANGE THE READING
CC FRAME AND RESULT IN INTRODUCTION OF STOP CODONS DOWNSTREAM OF THE
CC REPEAT REGION. THIS MAY BE A MECHANISM OF REGULATION AND A WAY TO
CC AVOID THE IMMUNOLOGICAL RESPONSE OF THE HOST (BY SIMILARITY).
CC -!- SIMILARITY: BELONGS TO THE TONB-DEPENDENT RECEPTOR FAMILY;
CC HEMOGLOBIN/HAPTOGLOBIN BINDING PROTEIN SUBFAMILY.
CC -!- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THE SEQUENCE WAS
CC ELONGATED IN THE N-TERMINAL SECTION, THEN A FRAMESHIFT WAS
CC INTRODUCED IN THE REPEATS REGION TO MAXIMIZE THE SIMILARITY WITH
CC OTHER ORTHOLOGS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32746; AAC22294.1; ALT_SEQ.
DR TIGR: HI0635;
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 2.
DR PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
KW Outer membrane; Transport; TonB box; Multigene family; Signal;
KW Receptor; Repeat; Complete proteome.
FT SIGNAL 1 24 POTENTIAL.
FT CHAIN 25 1063 PROBABLE HEMOGLOBIN AND HEMOGLOBIN-
FT HAPTOGLOBIN BINDING PROTEIN 1.
FT DOMAIN 26 53 7 X 4 AA TANDEM REPEATS OF Q-P-T-N.
FT REPEAT 26 29 1.
FT REPEAT 30 33 2.
FT REPEAT 34 37 3.
FT REPEAT 38 41 4.
FT REPEAT 42 45 5.
FT REPEAT 46 49 6.
FT REPEAT 50 53 7.
FT SITE 63 70 TONB BOX.
FT SITE 1046 1063 TONB C-TERMINAL BOX.
SQ SEQUENCE 1063 AA; 121160 MW; 370CB515523F2788 CRC64;
```

Query Match 4.7%; Score 229; DB 1; Length 1063;  
Best Local Similarity 19.0%; Pred. No. 7.6e-07;  
Matches 211; Conservative 152; Mismatches 402; Indels 344; Gaps 47;

```
QY 41 EDVHVAKRVPKDKVFTDARAVSTRODIEKSSNLDNIYRSIPGFTQODKSSGIVSLN 100
Db 72 ENINVKEKKVGETO-----ISAKKIAQO-----ASDRD-LVRYETGTVVETGRTGASGYA 123
QY 101 IRGDSGGRVNTVDGITQTFYSTSDAG-----RAGSSQFCASVDSNFIAGLDVKVKG 154
Db 124 VRGDE-NRVGINVDGLRQA--ETLSQGFKELFEGVGNFNTRNSIEIENVKTATITKG 180
QY 155 SFGSAGINSLAGSNRTLGDDVVOGNTYGLLLKGLTGTNSTKGNAMAAIGARKWLE 214
Db 181 ADSLKSAGLGGSVIPETKDARDYLDKDYLSYKRGYQTMNQNLKTLTLAGRSKKFD 240
QY 215 SGASGVGLYGH-----RRSVAQNYRVGGGQGHGNEGAELYER 253
Db 241 ILIIDTRDGHENIYDKIYPNKAQDLRAVGPTRKADPYQITRQSTLI-KLGFQPNEN 299
QY 254 RKQRYFQEGGLKFNSSNGKWERDFQRYWKTKYQKYNDPOELQKYIEGHDKSWRENLA 313
Db 300 HRLSVALDDSTLETKGIDLSYA--LRPY-STANNEXYG-----ERLI--NDQSKRKNIQ 348
QY 314 POYDITPIDSSLSKQOAGNLFKLEYDGVFNKYTAQPRDLNTKIGSKRIIN--RNVQFNY 371
Db 349 FSVE-----NFSQTPFWDHILKLSY-----SSOKITNKARSDEYCH 383
QY 372 GLSLNVAN--LNLTAAYNSGRQKYPKG---SKFTGWGLLKDFETYNNAKILTD----- 419
Db 384 QSTCNGVSNPQGLHVEKGVIKIDKYGELESKEIGWS-----HEFKNSKGEDADKDI 438
QY 420 -----LNNATFRLPRETELETQTLGNYFHNE-----YKG----- 449
Db 439 SQRSLDSVLINCEKDCSKKFRLYQYEDNSSEKYYDDRELEVGLTPNGKKGKPLK 498
QY 450 -----NRPEELGLFDPG-----DOD----- 466
Db 499 KGKTPSNWGPQETARELPFKSYGYSTDFVNDRLNTHTOQIKLDLKEFLHWTQHOLK 558
QY 467 -NGLYS-----YLGFRKDGKLLLPQKST---IVOPAGSQYF 498
Db 559 YGGLYEKTLKSMVNHQYNTAANVQWADYFFCARAKG--GNLGEKTPHPNVSVAGCVNG 616
QY 499 NTFVFAALKKDIYRLNSTNT-----VGVRF----- 525
Db 617 TPLHSD--IGKDYLIPIVTKNNVLYPGDNVQLTSWGLDLNLYRDHVKYLPGYDEKTPV 674
QY 526 -GGEYTYGSDDEKFRAGFENSPYKHKNCQSGIYEP-----VLKKYKKRRANNH 576
Db 675 PGGLIAGIFVPFNEKGVVYAYVPSGYKDCRYNTECYKKNFEENLALLRKTQYKH-HSY 733
QY 577 SVSISADFGYFMPFASYSRTHMPNIOEMY-----FSQIGDSGVHTALKPERANTWQ 629
Db 734 NLGLNLDPTDLWRVOLKRYANAFRAPSDEIYMTFKHPDFS-----IGPNTNLKAETAKTE 789
QY 630 FGFTYRKGLLKQDDTLGLKLVGYRSRIDNYIHNVYKQWMDLNGNIPSWVSSTGLAYTIQ 689
Db 790 VAFTFYKENSY-----LTLSAFQSDYRNFIDLVEFK-----NKQID--KGSIAIEPFY 835
QY 690 HRNFKKHVHKGFELELNYOYGRFTNL-----SYAYKSTQPTNFSDAESPNNASK 742
Db 836 QNQRNDAQRVGIEIASRLEMGLFEKLOGFHLGYKLTYOK-----GRIK 880
QY 743 EDQLKQCYG-----LSRVSAIPRDYGRLEVTGRWLNKLTILG--GAMRYFGKSTRAT 792
Db 881 DNKLRSYAEFLKLNPOYTAIASODQPMNALQPTT-----SVYINIGYDAPSKKMGMDYIIT 936
QY 793 --ABERYIDGTNGCNTSNVR-----OLGKRSIKOTETLAR--OPLIFDFY 833
Db 937 DVAAKKADSFNSQWTSWVRKENIYERTVPAQANGKVDKDSRGLWNRNRYTVIDTI 996
QY 834 AAYEPKKNLIFRAEVKNLFDORRYIDPLDAGNDAATQRYYSFDPKDKDEEYVTCNADKTL 893
Db 997 AYWKPIKNLFTAGVYNLTKKYL---TWDSARSVRHLG-----TINRVET-- 1039
QY 894 NGKYGGTSKSVLTFNFRGRFTLITMSYKF 922
```

```
Db 1040 -----ATGKGLNRFYAPGRNRYMSVQFEF 1063
RESULT 13
LBPA_NEIMA
ID LBPA_NEIMA STANDARD; PRT; 944 AA.
AC Q9JTK4;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DE 01-MAR-2002 (Rel. 41, Last annotation update)
DE Lactoferrin binding protein A precursor.
GN LBPA OR NMA1739.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=20222556; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
CC -!- FUNCTION: UNKNOWN. MAY BE AN IRON-SIDEROPHORE RECEPTOR.
CC -!- SUBCELLULAR LOCATION: Outer membrane.
CC -!- SIMILARITY: LOCAL TO OTHER TONB-DEPENDENT RECEPTOR PROTEINS.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC EMBL: AL162757; CAB84967.1; -.
CC InterPro: IPR000531; TonB_boxC.
CC Pfam: PF00593; TonB_boxC; 1.
CC PROSITE: PS00430; TONB_DEPENDENT_REC_1; FALSE_NEG.
CC PROSITE: PS01156; TONB_DEPENDENT_REC_2; 1.
CC Outer membrane; Iron transport; Transport; TonB box; Signal; Receptor;
CC Complete proteome.
CC SIGNAL 1 27 POTENTIAL.
CC CHAIN 28 944 LACTOFERRIN BINDING PROTEIN A.
CC SITE 827 944 TONB C-TERMINAL BOX.
CC SEQUENCE 944 AA; 105732 MW; CE06B6192E74AE3E CRC64;
Query Match 4.6%; Score 226.5; DB 1; Length 944;
Best Local Similarity 19.5%; Pred. No. 9.4e-07;
Matches 210; Conservative 155; Mismatches 379; Indels 335; Gaps 52;
QY 23 SYADAGRAGSE-AQIOVLEDVHVAKRVPKDKKVFVDARAVSTRODIEKSSNLD---- 77
Db 22 SYANPETASDAQAQSLSKEITVRAAKVGRSKEATGL-----GKIVTSETLNKEQV 75
QY 78 ----NIVRSIPGAFQODKSSGIVSLNIRGDSGFRVNTWVDGITQ-----TFYSTSDAG 129
Db 76 LGIRDLTRYDPGVAVVEQNGCAGSGYSIRGVDK-NRVAVSVDGVAQIAFTVQGSLSGYG 134
QY 130 RAGSSQFGASVDSNFNTAGLDVVKGSFGSAGINSLAGSNLRTLGDDVVOGNTYGLLL 189
Db 135 GRGGSGAIN-EIEYENISTVEIDKAGSDHSGALGAGAVAFRTKEADLSDGKSWGIG 193
QY 190 LKGIJGTNTSKGNAMAAIGARKWLESGASGVGLYGHRSVAQNYRVGGGQGHGNEGA 249
Db 194 AK--TAYGSKNRQPMKSLGAGFSKDGWEGLLIRTEROGRPHGDIADGVE-----Y 247
```





```
Db 615 KKDSEALSNIAA-----SKIGVGVNVALVKDKFTVYCATVTHYAAQ----- 656
Qy 802 NGGNTSNVRQLGKRRIKOTETLARQPLIFDFYAAYPEK----KNLIFRAEVKNLFDRIYI 857
Db 657 -----RRVPKDHVTPSYILTDLRATYAPLKGWKNLRLDFALENLFDRIYQ 704
Qy 858 DPLD-----AGNDAATQRIYS 873
Db 705 PAFSLMEGTGRNAKISAVYS 724
```

Search completed: July 24, 2002, 09:00:10  
Job time: 401 sec

GenCore version 4.5  
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: July 24, 2002, 08:52:04 ; Search time 68.49 seconds  
(without alignments)  
2328.826 Million cell

Title: US-09-762-926-4  
 Perfect score: 4904  
 Sequence: 1 MRSSFELKPICFYLMGVMLY.....SVLTNFARGRTFLITMSYKF 922

Scoring table: BLOSUM62  
Gapop 10.0 , Gapext 0.5

Searched: 562222 seqs, 172994929 residues

Total number of hits satisfying chosen parameters: 562222

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 s

```
Database :
SPTRMBL_19:*
1:  sp_archaea:*
2:  sp_bacteria:*
3:  sp_fungi:*
4:  sp_human:*
5:  sp_invertebrate:*
6:  sp_mammal:*
7:  sp_mhc:*
8:  sp_organelle:*
9:  sp_phase:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_invertebrate:*
14: sp_unclassified:*
15: sp_tvirus:*
16: sp_bacteriap:*
17: sp_archaeap:*
```

## SUMMARIES

Result No.	Query %			DB	ID	Description
	Score	Match	Length			
1	4840	98.7	922	16	Q9JTN4	Q9Jtn4 neisseria m
2	4768.5	97.2	921	16	Q9JR50	Q9Jr50 neisseria m
3	1085	22.1	755	16	Q9PIV1	Q9piv1 campylobact
4	903	18.4	753	2	Q93Q32	Q93q32 shigella fl
5	902	18.4	753	2	Q933S4	Q933s4 escherichia
6	459	9.4	885	16	Q92NX4	Q92nx4 rhizobium m
7	425	8.7	883	2	Q9X5P4	Q9x5p4 pseudomonas
8	425	8.7	891	16	Q9HYJ7	Q9hyj7 pseudomonas
9	425	8.7	989	16	Q9HUR6	Q9hur6 pseudomonas
10	415	8.5	916	2	Q9RHT4	Q9rht4 pseudomonas
11	396.5	8.1	899	2	P72473	P72473 serratia ma
12	351	7.2	848	16	Q9CKJ4	Q9ckj4 pasteurella
13	339.5	6.9	747	16	Q98L70	Q98l70 rhizobium l
14	312.5	6.4	782	2	Q93SH4	Q93sh4 bradyrhizob
15	308	6.3	784	16	Q9K1W6	Q9klw6 vibrio chol
16	278.5	5.7	739	2	O87381	O87381 haemophilus

## ALIGNMENTS

```

RESULT      1
Q9JTN4      PRELIMINARY;          PRT;       922 AA.
ID           AC Q9JTN4;
DT           01-OCT-2000 (T-EMBLrel. 15, Created)
DT           01-OCT-2000 (T-EMBLrel. 15, Last sequence update)
DT           01-DEC-2001 (T-EMBLrel. 19, Last annotation update)
DE           PUTATIVE OUTER MEMBRANE SUBSTRATE BINDING PROTEIN.
GN           NMA1700.
OS           Neisseria meningitidis (serogroup A).
OC           Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX           NCBI_TaxID=656699;
RN           [1]
RP           SEQUENCE FROM N.A.
RC           STRAIN=Z2491 / SEROGROUP A / SEROTYPE 4A;
RX           MEDLINE=20222556; PubMed=10761919;
RA           Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA           Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA           Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA           Jagels K., Leather S., Moule S., Mungall K., Quail M.A.,
RA           Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA           Whitehead S., Spratt B.G., Barrall B.G.;
RT           "Complete DNA sequence of a serogroup A strain of Neisseria
RL           meningitidis Z2491."
RL           Nature 404:502-506(2000)..
DR           EMBL; ALI62756; CAB84928.1; -.
DR           InterPro; IPR000531; TonB_boxC.
DK           PROSITE; PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW           Complete proteome.
SQ           SEQUENCE    922 AA;   104290 MW;   B45E683985DE20B6 CRC64;

Query Match              98.7%; Score 4840; DB 16; Length 922;
Best Local Similarity    98.7%; Pred. No. 3, 9e-260;
Matches 910; Conservative 3; Mismatches 9; Indels 0; Gaps 0;

Oy 1 MRSSFRLKPTICFYLMGVMLYHHSYAEADAGRAGSEAOIQVLEDVHVKAARVPDKKVFETA 60
Db 1 MRSSFRLKPTICFYLMGVMLYHHSYAEADAGRAGSEAOIQVLEDVHVKAARVPDKKVFETA 60

```

QY	61	RAVSTRQDIFKSSNLNIVRSIPGFTQDKSSGIYSLNIRGDSGFRVNTWVDGITQT	120
Db	61	RAVSTRQDIFKSSNLNIVRSIPGFTQDKSSGIYSLNIRGDSGFRVNTWVDGITQT	120
QY	121	FYSTSTADAGAGSSQFGASVDSNFTAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV	180
Db	121	FYSTSTADAGAGSSQFGASVDSNFTAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV	180
QY	181	QGNNTYGLLLKGLTGNTSTGNMAAIGARKWLESASGVLYGHSRVAQNYRVGGGG	240
Db	181	QGNNTYGLLLKGLTGNTSTGNMAAIGARKWLESASGVLYGHSRVAQNYRVGGGG	240
QY	241	QHIGNFGAEYLERRKQRYFVQEGGLKFNSSGKWERDFQRPYWKTKWYKYNPDQELQKY	300
Db	241	QHIGNFGAEYLERRKQRYFVQEGGLKFNSSGKWERDFQRPYWKTKWYKYNPDQELQKY	300
QY	301	IEGHDKSWRENLAPOYDITPIDPSSLKQOQSAGNLFKLEYDGVNKNKYTAQFRDLNLTIGSR	360
Db	301	IEGHDKSWRENLAPOYDITPIDPSSLKQOQSAGNLFKLEYDGVNKNKYTAQFRDLNLTIGSR	360
QY	361	KIINRNYQFNGLSLNSYANLNTAAYNSGROKYPKSGKETGWLKDKDFETYNNAKLIDL	420
Db	361	KIINRNYQFNGLSLNPYTNLNTAAYNSGROKYPKSGKETGWLKDKDFETYNNAKLIDL	420
QY	421	NNATFRLPRETELOTTLGFNFHNEYGKNNRFPPEELGDFDQDQNGLYSYLGRFKGDK	480
Db	421	NNATFRLPRETELOTTLGFNFHNEYGKNNRFPPEELGDFDQDQNGLYSYLGRFKGDK	480
QY	481	GLLPQKSTIVQAGSQYFNTFYDAALKDIIYRLNSTNTVGYRFGGEYTGYYGSDDEFK	540
Db	481	GLLPQKSTIVQAGSQYFNTFYDAALKDIIYRLNSTNTVGYRFGGEYTGYYGSDDEFK	540
QY	541	RAFGENSPYKHKCNQSCGIEYPLVKYKGRANNHNSVLSADFGDYFMPFASYSRTHRM	600
Db	541	RAFGENSPYKHKCNQSCGIEYPLVKYKGRANNHNSVLSADFGDYFMPFASYSRTHRM	600
QY	601	PNIOEMYFSQIGSGVHTALKPERANTWQGFNTYKGLLKQDDTLGLKLGVYRSRIDNY	660
Db	601	PNIOEMYFSQIGSGVHTALKPERANTWQGFNTYKGLLKQDDTLGLKLGVYRSRIDNY	660
QY	661	IHNVYKQWDLNGLNPSWVSTGLAYTIOHRNFKDKVHKHGFLELNDYDGRFFTNLSYA	720
Db	661	IHNVYKQWDLNGLNPSWVSTGLAYTIOHRNFKDKVHKHGFLELNDYDGRFFTNLSYA	720
QY	721	YQKSTQPTNFSADSESPNNAKEDQLKQYGLSRVSALPRDYGRLEVGTRWLNKLTLLG	780
Db	721	YQKSTQPTNFSADSESPNNAKEDQLKQYGLSRVSALPRDYGRLEVGTRWLNKLTLLG	780
QY	781	AMRYFGKSIRATAERYIDCTNGNTSNVRLGKRSIKQETTLARQPLIFDFVAAYEPKK	840
Db	781	AMRYFGKSIRATAERYIDCTNGNTSNVRLGKRSIKQETTLARQPLIFDFVAAYEPKK	840
QY	841	NLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEVTCKNADTLNCGYGGT	900
Db	841	NLIFRAEVKNLFDRIYDPLDAGNDAATQRYYSFDPKDKDEVTCKNADTLNCGYGGT	900
QY	901	SKSVLTNFARGRTFLITMSYKF 922	
Db	901	SKSVLTNFARGRTFLITMSYKF 922	
RESULT	2		
QJR50			
ID	QJR50	PRELIMINARY; PRT; 921 AA.	
AC	QJR50:		
DT	01-OCT-2000	(T=EMBLrel. 15, Created)	
DT	01-OCT-2000	(T=EMBLrel. 15, Last sequence update)	
DE	01-DEC-2001	(T=EMBLrel. 19, Last annotation update)	
DE	PUTATIVE TONB-DEPENDENT OUTER MEMBRANE RECEPTOR (TONB-DEPENDENT RECEPTOR).		
GN	TDFH OR NMB1497.		
OS	Neisseria meningitidis, and		

OS	Neisseria meningitidis (serogroup B).		
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.		
OX	NCBI_TaxID=487, 491;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=IR1074;		
RA	Turner P.C., Thomas C.E., Stojiljkovic I., Elkins C., Kizel G.,		
RA	Ala Aldsen D.A., Sparling F.P.;		
RT	"Identification and Characterization of Putative TonB-dependent Outer		
RT	Membrane Proteins among Pathogenic Neisseriae.";		
RL	Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.		
RN	[2]		
RP	SEQUENCE FROM N.A.		
RC	STRAIN=MC58 / SEROGROUP B;		
RX	MEDLINE=20175755; PubMed=10710307;		
RA	Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,		
RA	Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,		
RA	Nelson W.C., Winn M.L., DeBoy R., Peterson J.D., Hickey E.K.,		
RA	Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,		
RA	Mason T., Ciecko A., Parksey D.S., Blair E., Ciftone H., Clark E.B.,		
RA	Cotton M.D., Uterback T.R., Khouri H., Qin H., Vamathevan J.,		
RA	Gill J., Scarlato V., Masignani V., Pizza M., Grandi G., Sun L.,		
RA	Smith J.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;		
RT	"Complete genome sequence of Neisseria meningitidis serogroup B strain		
RT	MC58.";		
RL	Science 287:1809-1815(2000).		
DR	EMBL: AF227418; AAF73907.1; -		
DR	EMBL: AE002499; AAF41853.1; -		
DR	TIGR: NMB1497; -		
DR	InterPro: IPR000531; TonB_boxC.		
DR	PROSITE: PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.		
KW	Receptor; Complete proteome.		
SQ	SEQUENCE 921 AA; 104222 MW; 67985B75EDF8819D CRC64;		
Query Match	97.2%; Score 4768.5; DB 16; Length 921;		
Best Local Similarity	97.7%; Pred. No. 3.6e-256;		
Matches	901; Conservative 8; Mismatches 12; Indels 1; Gaps 1;		
QY	1	MRSSFRLPKICFYLMGVMLYHSHVAEDAGRAGSEAQIQVLEDVHVAKRVPKDKVFTDA	60
Db	1	MRSSFRLPKICFYLMGVMLYHSHVAEDAGRAGSEAQIQVLEDVHVAKRVPKDKVFTDA	60
QY	61	RAYSTRQDIFKSSNLNIVRSIPGFTQDKSSGIYSLNIRGDSGFRVNTWVDGITQT	120
Db	61	RAYSTRQDIFKSSNLNIVRSIPGFTQDKSSGIYSLNIRGDSGFRVNTWVDGITQT	120
QY	121	FYSTSTADAGAGSSQFGASVDSNFTAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV	180
Db	121	FYSTSTADAGAGSSQFGASVDSNFTAGLDVVKGSFGSAGINSLAGSANLRTLGVDVV	180
QY	181	QGNNTYGLLLKGLTGNTSTGNMAAIGARKWLESASGVLYGHSRVAQNYRVGGGG	240
Db	181	QGNNTYGLLLKGLTGNTSTGNMAAIGARKWLESASGVLYGHSRVAQNYRVGGGG	240
QY	241	QHIGNFGAEYLERRKQRYFVQEGGLKFNSSGKWERDFQRPYWKTKWYKYNPDQELQKY	300
Db	241	QHIGNFGAEYLERRKQRYFVQEGGLKFNSSGKWERDFQRPYWKTKWYKYNPDQELQKY	299
QY	301	IEGHDKSWRENLAPOYDITPIDPSSLKQOQSAGNLFKLEYDGVNKNKYTAQFRDLNLTIGSR	360
Db	300	IEGHDKSWRENLAPOYDITPIDPSSLKQOQSAGNLFKLEYDGVNKNKYTAQFRDLNLTIGSR	359
QY	361	KIINRNYQFNGLSLNSYANLNTAAYNSGROKYPKSGKETGWLKDKDFETYNNAKLIDL	420
Db	360	KIINRNYQFNGLSLNPYTNLNTAAYNSGROKYPKSGKETGWLKDKDFETYNNAKLIDL	419
QY	421	NNATFRLPRETELOTTLGFNFHNEYGKNNRFPPEELGDFDQDQNGLYSYLGRFKGDK	480
Db	420	NNATFRLPRETELOTTLGFNFHNEYGKNNRFPPEELGDFDQDQNGLYSYLGRFKGDK	479
QY	481	GLLPQKSTIVQAGSQYFNTFYDAALKDIIYRLNSTNTVGYRFGGEYTGYYGSDDEFK	540





```
RP SEQUENCE FROM N.A.
RX MEDLINE=21289081; PubMed=11395459;
RA SmaJ's D., Weinstock G.M.;
RT "The Iron- and Temperature-Regulated cjrBC Genes of Shigella and
RT Enteroinvasive Escherichia coli Strains Code for Colicin Js Uptake.";
RL J. Bacteriol. 183:3958-3966(2001).
DR EMBL; AF283294; AAK67309.1; -.
SQ SEQUENCE 753 AA; 82823 MW; 493A3B656FACA33F CRC64;

Query Match      18.4%; Score 903; DB 2; Length 753;
Best Local Similarity 28.9%; Pred. No. 4.7e-42;
Matches 274; Conservative 129; Mismatches 298; Indels 248; Gaps 32;

QY 16 GVML-----YHHSYAE--DAGRAGSEAQIOVLEVDVHVHAKRVPKDKKVFYTDARA VSTQD 68
DB 11 GILLSCGAYSQSISEKTNDSKGA-AEFSPLS---VSVGKTTSEQALEKTGATSSR-T 65
QY 69 IFKSSNLNIVRIPGAFQDQKSSGIVSLNIRGDSGFRVNTMVDGITQTFYSTDA 128
DB 66 TDKNLQSLDATVRSPGTYTQIDPGGAGISVNIIRGMSGFRVNTMVDGITQSFYGTSTG 125
QY 129 GRAGGS--SOFGASVDSNFTAGLDVVKGSFGSAGINSLAGSANLRLTGLVDVVQGNNTY 186
DB 126 TTAGHSTNNMAGVLIDPNLLVAVDVTGRDSSGSEGINALAGSANMRTIGVDDVIFNGNTY 185
QY 187 GLLLKGLTGTNKTGNAMAAIGARK--WLESASGVLYGHSSRVAQNTYRVGGGGHIG 244
DB 186 GLRSRFSVSGNGLGRSGMIALGGKSDAFTDTG-SIGYMAAVSGSSVYSNFSNGSG----- 239
QY 245 NFGAEYLERRKQRYFVQEGGLKFNNSGKWERDFQRPYWKTKWYKYNDFPOELQKYEIGH 304
DB 240 -----INSKEFGYDK-----YMKONPKSOLYK----- 261
QY 305 DKSWRENLAPOYDITPDPSSLKQOASAGNLFKLEYDGVFNKYTAQFRLNATKICSRKIIN 364
DB 262 -----MDIRPDE-----FNSFELSARTYENKFTRRDITS 290
QY 365 RNYQFNGLS-LNSYANLNLTAAYNSGRQKYPKGSFETGMLLKDFETYNNAKILDNLNT 423
DB 291 DDYIYKHYPFSELDIFNVTASTSRGNQKRYDGSLYTFY----KTSQAQRSDALDINNT 346
QY 424 ATFLPRETELQTTGLGNFYHNEYGNRPFPEELGFLPDGPDQDNGLYSYLGRFKGDKGLL 483
DB 347 SRTTV-ADNELEFMLGSKLMRTY-----DRTIHSAAAGDPKANOESI 387
QY 484 POKSTIVQAGSOYFNTFYDAALKDI-----YRLNVTNTVGYRFGEGYTYGSDDEF 539
DB 388 --ENNFPAPSGOODISALYTGKLVTRGIWEADFNLTNRNI-----TG----- 429
QY 540 KRAFGENSPYKKHCNQS--C---GIYEPVLKKGKRRANNHVSISADSGDFMFPFASY 594
DB 579 --IIRQ--DLPOIPTIIDLYGTVEYNLTLKLSVQNLNMRDYSEALNKLN----- 725
```

```
QY 874 SFDPKDKDEEVTGNADKTLNCGKYGGTSKSVLTNFAARGTRFLITMSYKF 922
DB 726 -MMPGLGDESH PANS-----ARGTWIFGGDIRF 753

RESULT 5
Q933S4 PRELIMINARY; PRT; 753 AA.
ID Q933S4
AC Q933S4;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-DEC-2001 (TReMBLrel. 19, Last annotation update)
DE CJRC.
GN CJRC.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21289081; PubMed=11395459;
RA SmaJ's D., Weinstock G.M.;
RT "The Iron- and Temperature-Regulated cjrBC Genes of Shigella and
RT Enteroinvasive Escherichia coli Strains Code for Colicin Js uptake.";
RL J. Bacteriol. 183:3958-3966(2001).
DR EMBL; AF283293; AAK67307.1; -.
DR EMBL; AF283288; AAK67303.1; -.
SQ SEQUENCE 753 AA; 82837 MW; 493A3B606FFCA33F CRC64;

Query Match      18.4%; Score 902; DB 2; Length 753;
Best Local Similarity 28.9%; Pred. No. 4.7e-42;
Matches 274; Conservative 129; Mismatches 298; Indels 248; Gaps 32;

QY 16 GVML-----YHHSYAE--DAGRAGSEAQIOVLEVDVHVHAKRVPKDKKVFYTDARA VSTQD 68
DB 11 GILLSCGAYSQSISEKTNDSKGA-AEFSPLS---VSVGKTTSEQALEKTGATSSR-T 65
QY 69 IFKSSNLNIVRIPGAFQDQKSSGIVSLNIRGDSGFRVNTMVDGITQTFYSTDA 128
DB 66 TDKNLQSLDATVRSPGTYTQIDPGGAGISVNIIRGMSGFRVNTMVDGITQSFYGTSTG 125
QY 129 GRAGGS--SOFGASVDSNFTAGLDVVKGSFGSAGINSLAGSANLRLTGLVDVVQGNNTY 186
DB 126 TTAGHSTNNMAGVLIDPNLLVAVDVTGRDSSGSEGINALAGSANMRTIGVDDVIFNGNTY 185
QY 187 GLLLKGLTGTNKTGNAMAAIGARK--WLESASGVLYGHSSRVAQNTYRVGGGGHIG 244
DB 186 GLRSRFSVSGNGLGRSGMIALGGKSDAFTDTG-SIGYMAAVSGSSVYSNFSNGSG----- 239
QY 245 NFGAEYLERRKQRYFVQEGGLKFNNSGKWERDFQRPYWKTKWYKYNDFPOELQKYEIGH 304
DB 240 -----INSKEFGYDK-----YMKONPKSOLYK----- 261
QY 305 DKSWRENLAPOYDITPDPSSLKQOASAGNLFKLEYDGVFNKYTAQFRLNATKICSRKIIN 364
DB 262 -----MDIRPDE-----FNSFELSARTYENKFTRRDITS 290
QY 365 RNYQFNGLS-LNSYANLNLTAAYNSGRQKYPKGSFETGMLLKDFETYNNAKILDNLNT 423
DB 291 DDYIYKHYPFSELDIFNVTASTSRGNQKRYDGSLYTFY----KTSQAQRSDALDINNT 346
QY 424 ATFLPRETELQTTGLGNFYHNEYGNRPFPEELGFLPDGPDQDNGLYSYLGRFKGDKGLL 483
DB 347 SRTTV-ADNELEFMLGSKLMRTY-----DRTIHSAAAGDPKANOESI 387
QY 484 POKSTIVQAGSOYFNTFYDAALKDI-----YRLNVTNTVGYRFGEGYTYGSDDEF 539
DB 388 --ENNFPAPSGOODISALYTGKLVTRGIWEADFNLTNRNI-----TG----- 429
QY 540 KRAFGENSPYKKHCNQS--C---GIYEPVLKKGKRRANNHVSISADSGDFMFPFASY 594
DB 579 --IIRQ--DLPOIPTIIDLYGTVEYNLTLKLSVQNLNMRDYSEALNKLN----- 725
```

```

Db 430 -----YKACDSRVICVQGSYDIDKEGG-----FNPSVQLSAQVTPWLQPIGY 475
QY 595 SRTHMPNIQEMTFSGIDSGVITALKPERANTWQFGFNTYKGLLKQDDTLGLKLVGYR 654
Db 476 SKMRAFNIEQEMTFSGGASMPFLKPERAETWQAGFNITDRLLVEODALRFKALAYR 535
QY 655 SRIDNYIHN-----VYKWKW-----DLNGNIPSWSSSTGLAYTIQHRNF 693
Db 536 SRQNYIYSYSLVCSGRKCSMAEVIGNDWEGISDYSNNMIIYVNSA-----584
QY 694 KDKVHKHGFLELNLYDGRFTNLAYAKSTQPTNFSASESPNNASKEDQLKQGYGLS 753
Db 585 SDVIK -GFELEMDYDAGFAFGRLSFSQQTDOPTSTASY-----FGAG 628
QY 754 RVSAIPRDYGRLEVGTRWLNKNTLGLGAMRYFKSIRATAEERYIDGTNGNTSNVRQLG 813
Db 629 DMTLEPKRYMTLDTGVFFONALTGLTIKYGKARLSPDFEQDEHTGA-----678
QY 814 KRSTKQETLAROLPLIDFVAAYEPKKNLIFRAEVKNLFDORRYIDPLDAGNDAATQRYYS 873
Db 679 --IIKQ--DLPQPTIIDLYGTVEYNRLTLKLSVONLMNRDYSEALKNL-----725
QY 874 SFDPKDKDEEVTNADKTLCLNGKYGTSKSVLTNFARGRTFLITMSYK 922
Db 726 -MPEGLDETHPANS-----ARGRTWIFGGDIRF 753

RESULT 6
Q92NX4 ID Q92NX4 PRELIMINARY; PRT; 885 AA.
AC Q92NX4;
DT 01-DEC-2001 (TrEMBLrel. 19, Created)
DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
DE PUTATIVE IRON/HEME TRANSPORT PROTEIN.
GN SMC04205.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021.
RX MEDLINE=21368234; PubMed=11474104;
RA Galibert F., Finan T.M., Long S.R., Puehler A., Abola P., Ampe F.,
RA Barloy-Hubler F., Barnett M.J., Becker A., Boistard P., Bothe G.,
RA Boutry M., Bowser L., Buhrmester J., Cadieu E., Capela D., Chain P.,
RA Cowie A., Davis R.W., Dreano S., Federspiel N.A., Fisher R.F.,
RA Gloux S., Godrie T., Goffeau A., Golding B., Gouzy J., Gurjal M.,
RA Hernandez-Lucas I., Hong A., Huizar L., Hyman R.W., Jones T., Kahn D.,
RA Kahn M.L., Kalman S., Keating D.H., Kiss E., Komp C., Lelaure V.,
RA Masuy D., Palm C., Peck M.C., Pohl T.M., Portetelle D., Purnelle B.,
RA Ramsperger U., Surzycki R., Thebaud P., Vandenbol M.,
RA Vorholter F.J., Weidner S., Wells D.H., Wong K., Yeh K.-C., Batut J.;
RT "The composite genome of the legume symbiont Sinorhizobium meliloti.";
RL Science 293:668-672(2001).
DR EMBL; AL591789; CAC46612.1; -.
KW Complete proteome.
SQ SEQUENCE 885 AA; 95389 MW; BEE36641060DFBF4 CRC64;

```

```

Query Match 9.4%; Score 459; DB 16; Length 885;
Best Local Similarity 22.9%; Pred. No. 2.le-17;
Matches 224; Conservative 122; Mismatches 332; Indels 300; Gaps 42;

QY 25 AEDAGRAGSEAOIOLVEDVHVAKRVPKDK-----VFTDARAVS-TRQDIFKS- 72
Db 128 AGDAGPCEGEGATLVRIYVTKTRGNANSAGAGFGTDPWYEEPASVSVSRDAVQR 187
QY 73 -SENLDNIVRSIPGAFQDQSSGI-VSLNIRGDSGFRVNTMWDGITQTFYSTSTDAGR 130
Db 188 AARNANDVLDSVAGVTSNRSEAQNPGLAINVRGLQDNQRTTMDGARQDFQ-----R 240

```

```

QY 131 AGSSQFGASVDSNFIAGLDVVKGSFSGSAGINSLAGSANLRTLGVDVVGQNTYGLLL 190
Db 241 AGHGAQRVVYDTAFRLSRVEYKGAAGVAGAGSLGAVNPTVTTADITIPDRDRGVEL 300
QY 191 KGTTGTNSTKGNMAAIGARKWLSEASGVLYIGHSRSSVAQNYRVGGGG-----241
Db 301 NAETGTNAYFYNG-SLIGAARFSE--DFSVLGGISRRKVG-DYDFGONGKSPLLDLAVT 355
QY 242 -----HIGNFGAEYLERRKORYFVQEGGLKFNKSGKWERDQRPYWKTKYQKY 291
Db 356 TAVDDSLFSLRLETFTGL-----LKV-----EGSPSDDFT--FDLSWLR-- 392
QY 292 NDQELQKYIEGHDKSWRENLAPOYDITPDPSSLKQOSAGNLFKLEVDGVFNKYTAQFR 351
Db 393 NDSEAIQGGVLFGD--LRD-----DPQNYLNTVSSSEFWDPD-----SELI 432
QY 352 DLNATKIGSRKIIN--RNYQFNGLSLNSYANLNLTAAYNSGRQKYPKSGKFTGWLLKD 408
Db 433 DLKGLWYRVVNDLRYD-----PRLPITYAMTS-----FGG-----466
QY 409 FETYNNAKILDLNNTATFRLPRETELQTLGFNY---FHNEYKGNRPPEELGLFFDGPQ 465
Db 467 -----SLDNTSRF-----ETAL-GDLSLNYGGEAYSNGKTTTPPLV-----D 503
QY 466 DNGL---YSLGRFKGDKGLLPQKSTIVQAGSQYFNTFYFD-----AALKKD 510
Db 504 DQGFDEAYGYKG-----LNPVGRSRMTSAFLNATLEHDDWLVGAGLRD 548
QY 511 IYRLNSTNTVVG-----YRFGGEYTYGSDDEFKRAFGENSPYKXKHCNQCSCGIYEPVLK 566
Db 549 RYRLKGFTEVGGKPRYIVVPGVCYFYDDGEC--AYDEDPVY-----GGGEAVLE 598
QY 567 KYCKKRANN---HSVSISADFGDYFMPFASYSRTHRPNTQOEMYFSQIGDSGVHTA--- 619
Db 599 RVDIDKSGGALLPSARIAMVPFEGIQPFVYAHYRPPSVMEAL-----TSGGHPGDAIA 653
QY 620 -----LKERANTWQFGNTYKGLLKQDDTLGLKLVGRSRIDNYI--HNYYGKWD 670
Db 654 TYIPNPKLPERGRTWELGINARDGLFTAGDSLRLTVYVFDRTIQDYITLGNFYATFD 713
QY 671 LNGNIPSWSSSTGLAYTIQHRNFKDKVHKHGFLELNLYDGRFTNLNSYAYQKSTQPTNF 730
Db 714 KN-----LFQHVNLDDGDTMNGVEIEASYDMGSAVVGASYTYLKTIDYADTY 759
QY 731 S-----DASESPNNASKEDQLKQGYGRVSRVSRALPRDYGRLEVGTRWLNKNTLGLGAMRYFG 786
Db 760 SYSGPTASGTPLAASGNTVPV-----SVLFVPPENKFTLDAGIRLFEKRLVGLGRATY-- 812
QY 787 KSRATAEERYIDGTNGG--NTSNVRQLGKRSIKQETLARQPLIFDFAAYEPKKNLIF 844
Db 813 -----VSDSKPTVQLAGLFTAGYK-----VFDIYGSYSFSDSAKL 849
QY 845 RAEVKNLFDRIYIDPLDAGNDAATQRYYSFDPKDKDEEVTNADKTLCLNGKYGTSKSV 904
Db 850 RLAINVTDQYAPALCA-----FYYP-----871
QY 905 LTNFARGRTFLITMSYK 922
Db 872 ----APGRTATVSLNFKF 885

RESULT 7
Q9X5P4 ID Q9X5P4 PRELIMINARY; PRT; 883 AA.
AC Q9X5P4;
DT 01-NOV-1999 (TrEMBLrel. 12, Created)
DT 01-NOV-1999 (TrEMBLrel. 12, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE HEME RECEPTOR HASR.
GN HASR.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.

```

```
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA01;
RX MEDLINE=20121752; PubMed=10658665;
RA Ochsner U.A., Johnson Z., Vasil M.L.;
RT "Genetics and regulation of two distinct haem-uptake systems, phu and
RL has, in Pseudomonas aeruginosa.";
DR Microbiology 146:185-198(2000).
DR EMBL; AF127223; AAD31013.1; -.
DR InterPro; IPR00531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Receptor.
SQ SEQUENCE 883 AA; 96942 MW; A29B07A3A6C31DCF CRC64;

Query Match 8.7%; Score 425; DB 2; Length 883;
Best Local Similarity 22.7%; Pred. No. 1.6e-15;
Matches 215; Conservative 114; Mismatches 382; Indels 238; Gaps 37;

Qy 30 RAGSEAQIQVLEDVHVKAKRVPKDKKVFDTARAVST--RQDIFKS--SENLDNIVRSIPGA 86
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 117 RVSQDDLQVQSPSV--ISAAR--PDDWVYQTPHSVSVIGREQIERNPPRHAADMLEETPGV 173

Qy 87 FTQODKSSGIYSLNIRGDSGFRVNTWVDGITQTFTYSTDAGRAGSSQFGA-SVDSNF 145
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 174 YSSVSQDPGLSVNIRGIQDYGYNMSVDGMQRNYQ-----QSGHQRNGLTYVDP 226

Qy 146 IAGLDVVKGFSGSGAGINSLAGSANLRTLGVDVVQGNNTYG---LLLKGLTG-TNSTK- 200
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 227 LSEVVDKGSAMGAGVIGGIANFRTLEARDLVRPKQVGGVRVLTSLGGDANGTHF 286

Qy 201 -GNAMAAIGARKWLESGASGVLYGHSSRRSVAQNYRVGGGQGHIGNFGAEYLERRKORYE 259
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 287 IGSAAFAIGTEVW-----DMLVAASERHLG-DYDPGTGKS-IG-----ELRTGAWF 330

Qy 260 VOEGGLKFNNSGKWERDFORPYWTKYQYNDPQELQKIEGHDKSWRENLAPOVDIT 319
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 331 NPEAGORVK-----HSPVAYSGV-----MRSRLAKLGV 361

Qy 320 PIDPSSLKQOAGNLFKLEYGVFNKYTAQPRDLNTKIGSRKIINRNQYQNYGLSL-NSY 378
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 362 PQDQ---RLQFSYLTQVSYDDA-NMLNTENQALWEKLGSSDVAQNAFIDYGVAPDNPL 417

Qy 379 ANLNLTRAY--NSGRQYPKGSKFTGNGLLKDFETYNNAKILTDJNNTATRLPRETELQT 436
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 418 VDFRAKLYYVDNRNQOTLQGITPGYSITVTDTYG---AQAOQSTFALDLSLTRA 473

Qy 437 TLGFNYPHNEYGKNR-----FPEELGFFDGDQDNGLYSYLGRFKGDKGLLPQ 485
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 474 NYGLEFFYDKVRPDSQPRASTSAVGPPAAEGM---TPKGDRLGSLFARLDYD----- 524

Qy 486 KSTIVQAGSQYFNTFFDAALKDIIYRL-----NSTNTVGYRFGGEYTYGYSDEDF 539
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 525 -----YDDLNLNAGLRVDRYLRGDTGCFNARTILGTTRTDMPLQVAVD--- 570

Qy 540 KRAFGENSPYKHCNCSGIEYVPLKKYKKRANNSHSVISADFG-DYFMPFASYSRTH 598
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 571 -REEGRESPTF-----GLSVKPGVDMLQLFATYKGW 601

Qy 599 RMPNIQEWYSQIGDSG-----VHTALKPERANTWQFGENTYKGLLKQDDTLGLKLVG 652
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 602 RPAVATESLTJGRPHGGGAENMYNPNPLSPERSKAWVEGVFNVLKENLWFSDDRGLKVA 661

Qy 653 YRSRIDNVIHNVYKWNLDNGNTPSWSSSTGLAVTIQHRNFKDKVHKHGFELNLYDGR 712
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 662 FDTVDVDFIFMGGMQ-----PPGYCMAGIGNSAVNNL-DSTRPRGEYQLDYDAGL 713

Qy 713 FFTNLSYAYOKS-----TQPTNFSASESPNNASKEDQLKOGYGLSRV--- 755
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 714 AYGOLSYTHMIGNSDFCSKTAWLGGVTVTKVSGRRRPVDMRPEQANAATHCSAVLGS 773

Qy 756 -SALPRDYGRLGVGTRWLNKLLGGAMRYFGKSIRATAEERYIDGTNGGNTSNVROLGK 814
```

```
Db 774 AEHPMDRG-----SLTLG--MRFDRRLDVGARARYSEGYSVAGGATVSOAGV 820
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 815 RSIQQTETLARQPLIFDFYAAAYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAAATQRYSS 874
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 821 YPADWKEY-----TVYDLYGSYRVSDELTTLRLAMENVTDRAYLVPLG----- 862
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 875 FDPKDKDBEVTCTNADKTLCKNGYKGTSKSVLT-NFARGRTFLITMSYKF 922
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 863 -----DVLAFTLGRGRTLQGTLEYQF 883

RESULT 8
Q9HYJ7 PRELIMINARY; PRT; 891 AA.
AC Q9HYJ7;
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HEME ACQUISITION PROTEIN HASR.
GN HASR OR PA3408.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warriner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Saier M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen.";
RL Nature 406:959-964(2000).
DR EMBL; AE004762; AAG06796.1; -.
DR InterPro; IPR000531; TonB_boxC.
DR Pfam; PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 891 AA; 97864 MW; 126D6F41E7D66979 CRC64;

Query Match 8.7%; Score 425; DB 16; Length 891;
Best Local Similarity 22.7%; Pred. No. 1.6e-15;
Matches 215; Conservative 114; Mismatches 382; Indels 238; Gaps 37;

Qy 30 RAGSEAQIQVLEDVHVKAKRVPKDKKVFDTARAVST--RQDIFKS--SENLDNIVRSIPGA 86
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 125 RVSQDDLQVQSPSV--ISAAR--PDDWVYQTPHSVSVIGREQIERNPPRHAADMLEETPGV 181

Qy 87 FTQODKSSGIYSLNIRGDSGFRVNTWVDGITQTFTYSTDAGRAGSSQFGA-SVDSNF 145
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 182 YSSVSQDPGLSVNIRGIQDYGYNMSVDGMQRNYQ-----QSGHQRNGLTYVDP 234

Qy 146 IAGLDVVKGFSGSGAGINSLAGSANLRTLGVDVVQGNNTYG---LLLKGLTG-TNSTK- 200
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 235 LSEVVDKGSAMGAGVIGGIANFRTLEARDLVRPKQVGGVRVLTSLGGDANGTHF 294

Qy 201 -GNAMAAIGARKWLESGASGVLYGHSSRRSVAQNYRVGGGQGHIGNFGAEYLERRKORYE 259
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 295 IGSAAFAIGTEVW-----DMLVAASERHLG-DYDPGTGKS-IG-----ELRTGAWF 338

Qy 260 VOEGGLKFNNSGKWERDFORPYWTKYQYNDPQELQKIEGHDKSWRENLAPOVDIT 319
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 339 NPEAGORVK-----HSPVAYSGV-----MRSRLAKLGV 369

Qy 320 PIDPSSLKQOAGNLFKLEYGVFNKYTAQPRDLNTKIGSRKIINRNQYQNYGLSL-NSY 378
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 370 PQDQ---RLQFSYLTQVSYDDA-NMLNTENQALWEKLGSSDVAQNAFIDYGVAPDNPL 425
```

```
Qy 379 ANLNLTAA--NSGRQKYPKSGKTGWLKDKDFETYNNNAKILDLNNTATFPLPRETELQOT 436
Db 426 VDFKAKLYVDNRRQOTLQGITPGXSITQTDITY-----AQANTSTFALDLSILRA 481
Qy 437 TLGNFYHNEYGNK-----PPEELGLFFDGDQDNGLYSYLGRFGKDGKLLPQ 485
Db 482 NYGLEFFYDKVRPSSQPRASTSAVGPPAAAGM-----TPKGRALGSLFARLDYD----- 532
Qy 486 KSTIVQAGSOFNTFYDAAKKDIYRL-----NXTNTVGYRFGGEYTCYCSDDDEF 539
Db 533 -----YDDLNLNAGLRYDRYRLGDTGFNARTFELGTTRQDMPLOYAVD----- 578
Qy 540 KRAFGSPYTKKCNQSCGIIYEPVLKYYKKRANNSHVSISADFG--DYFMPFASYSRTH 598
Db 579 -REEGRFSPF-----GLSVKPGVDWLQLFATYKKGW 609
Qy 599 RMPNIOEYFQIGDSG-----VHTALKPERANTWQGFNTYKKGLLKQDDTLGLKLVG 652
Db 610 RPPAVTESLITGRPHGGAENWPNPFLSPERSKAWEGFNVLKENLWFSDDRLGLKAVY 669
Qy 653 YRSRIDNYIHNVYKGMWDLNIPSWSSSTGLAYTIQHRNEFKDKVHKHGFLELELNIDYGR 712
Db 670 FDRVDDDFIENGGMQ-----PPGYMGAGIGNSAYVNNL-DSTRFRGVEYQLDYDAGL 721
Qy 713 FFTNLISAYOKS-----TQPTNPSDASESPNNASKEDQLKQYGLSRV----- 755
Db 722 AYQGLSYTHMIGSNDFCSTKAWLGVTQTVKSGRRPVPIDMRPDEQANAATHCSAVLGS 781
Qy 756 -SALPRDYGRLEVGTRWLGKNTLGGAMRYFKGSIRATAEERYIDGTNGGNTSNVRLGK 814
Db 782 AEHPMDRG-----SLTGL--NRFFDRDLVGCARARYSEGISVAGCATVSQAGV 828
Qy 815 RSIKQTEFLARQPLIFDYAAEPKKNLIFRAEVKNLDRRIYIDPLDAGNDAATQRYSS 874
Db 829 YPADWKEY-----TVYDLYGSYRVSDELTLRLAMENVTDRAVLVPLG----- 870
Qy 875 FDPKDKDEVTCNADKTLCLNCKYGGTSSKVLN-FNARGTFLITMSYKF 922
Db 871 -----DVLAFTLGRGRTLOGTLEYQF 891

RESULT 9
Q9HUR6
AC Q9HUR6 PRELIMINARY; PRT: 989 AA.
DT 01-MAR-2001 (TrEMBLrel. 16, Created)
DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
DT 01-OCT-2001 (TrEMBLrel. 18, Last annotation update)
DE HYPOTHEICAL PROTEIN PA4897.
GN PA4897.
OS Pseudomonas aeruginosa.
OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
OC Pseudomonas.
OX NCBI_TaxID=287;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 15692 / PA01;
RX MEDLINE=20437337; PubMed=10984043;
RA Stover C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warrenner P.,
RA Hickey M.J., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
RA Garber R.L., Goltzy L., Tolentino E., Westbrook-Wadman S., Yuan Y.,
RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Larbig K., Lim R.M.,
RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
RA Reizer J., Sailer M.H., Hancock R.E.W., Lory S., Olson M.V.;
RT "Complete genome sequence of Pseudomonas aeruginosa PA01, an
RT opportunistic pathogen."
RL Nature 406:959-964(2000).
DR EMBL: AE004902; AAG08282.1; -.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC.1.
DR PROSITE: PS01156; TONB_DEPENDENT_REC_2; UNKNOWN_1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 989 AA; 108303 MW; 99D33D2FBD0F0806F CRC64;
```

```
Query Match 8.7%; Score 425; DB 16; Length 989;
Best Local Similarity 23.2%; Pred. No. 1.8e-15;
Matches 230; Conservative 130; Mismatches 382; Indels 250; Gaps 42;

Qy 30 RAGSEAQIQLVEDVHVAKRVPKDKVETDARAVSTRODIPKSSNLD-----NIVR 81
Db 149 RMSGEAPADLSPVVSAAELADPOKETYT-----APRSSVYLSSEDDIRFGRVSGDILLQ 203
Qy 82 SIFGAFTQODKSSGIVSLNIRGDSGFRVNTWVGITQTFYSTSTADAGRAGSSQFGASV 141
Db 204 GIPGVQVGDNRGALDNIIRGICQSVRVAVRVDGAEQ-----ALDVRVYAGTQOORSYI 258
Qy 142 DSNFTAGIDVYVKGFGSGAGINSLAGSANLRTLGVDDVVQGNNTYGLLLKGLTGTNSTKG 201
Db 259 DPLDVSSVTVDKGPSTRSGAI---GGSVEMRTIGVKDILVDGKDLGVRTGDVWNN--G 312
Qy 202 NAMAAGARKWLESGASV---GVLYGHSRR--SVAQYRVGGGQGHGNGFABYLERR 254
Db 313 VAPQHRASAKSTENLSSVPHDRGSLFGSAKSGSAFAFYR-----NEHL-DLVAAYAQRN 367
Qy 255 KORYFVQGGGLKFNNSGKWERDFQRPYK-----TKWY---QKYNPQELQKYE 302
Db 368 QGNFY-----SGKGDQRYRVNRYGRESSVAKYNAGEEVLNSSSETESL 415
Qy 303 GHDKSWRENLAPOYDIT-----PIDPSSLKQOAGNLKFLKLEYDGV-FNKYTAQF 350
Db 416 -LKATWR--IADEHTDLGYRRYDGTGEIMPSDIFREGTAGIYQYPLSEVKIDITYARY 472
Qy 351 RDLNTKIGSRKIIRNRYQFNGLSINYSANLNLTAAVNSGRQKYPKSGKFTWGLLKDFE 410
Db 473 RYLPEN-----NPLVDLSTGLWMTAKSMDLTSVLAPRSQAYRSRDNWT----- 516
Qy 411 TYNNAKI-LDLNNTATFRL-----PRETELOTLGFN-----YF 443
Db 517 RQDNRIIGDDLNNVARFETDFGDKLDLGGSFQVEDIOPKSVVYTLHDINARLRLDAT 576
Qy 444 HNEYGKN-----RPFEEGLFFDG-----PDQNGLYSYLGRFGKDKGLLPKQSTIVOPA 493
Db 577 ROEYGLNGKLEFKPVERLTLWGGGRIYSHFNKNDNGISAPRREDRDMRFI-----TVSRPG 632
Qy 494 -GSQYF---NTFYFDAA---LKKDIYRLAYSTNTVYRFGGEYTYGYSDDDEKFRAG 544
Db 633 YYGSMWVFPDQNGYTDATDPLNNGIVTNTNPNFEGIPF-----DEFGPA-- 679
Qy 545 ENSPTYKKHKQSCGIYEPVLKYYKKRANNSHVSISADF---GDYFMPFASYSRTHRP 601
Db 680 -NVTVHPSRVNTVYGYN--YSKKGSSRGGGFSFPAFGLNFELAPDTFV-YASYTEGLRLP 735
Qy 602 NIQEMYFSQIGDSGVHHTA--LKPERANTWQGFNTYKKGLLKQDDTLGLKLVGYSRIDN 659
Db 736 SLFE---TSQGTQLQVEPGKDLKSPERSWSWICASALRSLDADGSAAIKLAYFNNTIKN 792
Qy 660 YIHNVYKGMWDLNNGNIPSWVSTGLAYTIQHRNEFKDKVHKHGFLELYNVDYGRFFTNLSY 719
Db 793 YITRYD-----PGQMLMTFSNT-----DSTRTSGLQLSHYDAGRVEFADUSA 836
Qy 720 AYQKSTQPTNFSADSESPNNASKEDQLKQYGLSRVSL-----PRDYGRELVGTR 770
Db 837 TYLLKTETCDAAFAARLARAGANRYORTENTPNCPSFGWSYNTQNPPRLATNLTAGLR 896
Qy 771 WLGNKLTGGAMRYFGKSIIRATAEERYIDGTNGGNTSNVRLGKRSIKQTEFLARQPLIF 830
Db 897 FFDQALTLGGRMTY-TSGPTATADKPQWVG-----ATTPQIYERSVQ-----LF 939
Qy 831 DFYAAYEPKKNLIFRAEVKNLFDRIYIDPLDAGNDAATQRYSSFPDKDKDEVTCNADK 890
Db 940 DLFLKYLKLFTEHLNASLQNLTDRIYLDPL-----AQSPMP----- 975
Qy 891 TLCNGYGGTSGKSVLTNFAARGRTFLITMSYKF 922
Db 976 -----APGETLAVGQAKF 989
```





```
Db 745 PRLNLHATLGLTRPEQRLDYGARYSYSKRLVPVLSAERFV-----NTSSIE----- 791
QY 818 KQETLARQPLIFDYAYAEPPKKNLIFRAEVKNLFDRRYIDPLDAGNDAATQR 870
Db 792 -----WAPYSLV-DLYANYNVSNLKLTMMDNVFNRYIIDNNMGLNTAPGR 838

RESULT 13
Q98L70
ID Q98L70 PRELIMINARY; PRT; 747 AA.
AC Q98L70;
DT 01-OCT-2001 (Tremblrel. 18, Created)
DT 01-OCT-2001 (Tremblrel. 18, Last sequence update)
DE HEME ACQUISITION PROTEIN, HASR.
GN MRL155.
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneko T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti.";
RL DNA Res. 7:331-338(2000).
DR EMBL: AP002996; BAB48593.1; -.
DR InterPro: IPR000531; TonB_boxC.
DR Pfam: PF00593; TonB_boxC; 1.
KW Complete proteome.
SQ SEQUENCE 747 AA; 80102 MW; 4105DA001B73E752 CRC64;

Query Match 6.9%; Score 339.5; DB 16; Length 747;
Best Local Similarity 20.0%; Pred. No. 6.9e-11; Indels 341; Gaps 38;
Matches 188; Conservative 131; Mismatches 282;

QY 23 SYADAGRAGSEAIQVLEDVHVHAKRVPKDK-----KVFTDARAVSTRQDIFKS 72
Db 52 AHAAQATQPAGEQTDQ-----SKKADQEKATAGATLLDKILVISRTGETAIESLAS 103
QY 73 SENLD-----NIVRSIPGAFTQODKSSGGIVSLNIRGDSGFGGRVNTWVGITQ 119
Db 104 ASHVDQBLAHMATTPNEMLLGVPVATQADARRVSTSNIRGLQDFGRVAVIVDQARQ 163
QY 120 TFYSTSTDAGRAGSSQDFGASVDSNFTAGDVGKSGSGSAGINSLAGSANLRTLGVDDV 179
Db 164 DFQ-----RSDHGTOSTFYIDELVKVSDVIRGVPANTYGSAGGVVFFDTKDAADF 216
QY 180 VQGNNTYGLLLKGLTGNTSGNAMAIGAARKWLESQSYG-VLYGHRSRRSAQNTRVGG 238
Db 217 LKPEETWASVTGRYESNG-KGWTSATGAYRENNWMDALGNIVYRN-----YDNYK-DG 269
QY 239 GGQHIGNFGAEYLLERRKQRFVQEGGLKFNNSGKWERDQRPYWKTKYQKNDNQELQ 298
Db 270 GGDVNTGTGFDVL-----SGLLKTISI-----RP-----TENSELK 299
QY 299 KYIEGHDKSRENLAPOYDITPDPSLKQSQSAGNLFKLEYDGVFNKYTAQFRDLNLTIG 358
Db 300 LGWVGSSDGHDEYSGGM-----PVNDVDLKS-----NTFTA----- 330
QY 359 SRKIINRYQFNGLSUNYANLNLTAAYNYSGRQKYPKSGKFTGWGLLKDFETYNNAKIL 418
Db 331 -----RYNITDEKSLDLHINTSYN----- 351
QY 419 DLNN*ATFRPRETELOTTIGFNYFNEYKGNRFPPELGLFFDGPDDNGLSYLGRFKG 478
```

```
Db 352 -----KTNLDLT----- 358
QY 479 DKGLLPQKS-----TIVPAGSOYFNTFYFDDAALKK-DIYRLN-YSTNTVGYR--FGG 527
Db 359 --SLVPQKRFDPIITGLPTVLPAGSQ-----STFDVGTGTIDIMWTSRPFETGGIAHELTYGG 412
QY 528 EYTGYYGSDDEFK---RAFGEN--SPYKKKHQSCGIYE-----PVLKKYKGRANNNH 576
Db 413 DWVG-----DDVKTGGTAGDGSFYTPSGKKNVS---GAYVQDKLTWDNLEVIAGLRVDNY 464
QY 577 SV--SISADFGDYFMP-----FASYSRTHRMNPNIQEMYPFSGIDSG 615
Db 465 SLKDSRTSGDRLSPRITVGVSPFESAGLAGLQFYGTVAEGYRSPSLTETLISGNHPAG 524
QY 616 V-----HTALKPERANTWOGFNTYKGLLKQDDTLGLKLVGYSRIDNYIHNVYKWKW 669
Db 525 VTFPFLPNPNLRPTGKTEFGINRQNDIFEPGDALRVAAAYFHNNDVDDYIDGVTLSPF 584
QY 670 DLNNGNTPSMVSSSTGLAVTIQHRNFKDKVHKHGFELNLYDGRFFTNLSYAYQKSTQPTN 729
Db 585 APGSGCP---FGGIPICFOYQNA-QAKIDGFELEGVYDAGWGYAGLSAS----- 631
QY 730 FSDASESPNNASKEDQKQGYLS-----RVSALPRDYGRLEVGTRWLGKNTLGGAMR 783
Db 632 -----ITNGHTISYKGVADLATIPSSQVTAQGLRFLDKLTVGGEVQ 675
QY 784 YFGKSTRATAERYIDGTNGCGNTSNVRQLCKRSIKQTETLARQPLIFDFVAAAYEPKKNL- 842
Db 676 YNKGK-----KGN-----AVAEDYTLVNAFASYQATDNLK 705
QY 843 -IFRAEVKNLFDRRYIDPLDAGND-----AATQRY 871
Db 706 VDRAD--NLFVDVKYANPLNGSTTVAYVEPITLKAATMRF 745

RESULT 14
Q93SH4
ID Q93SH4 PRELIMINARY; PRT; 782 AA.
AC Q93SH4;
DT 01-DEC-2001 (Tremblrel. 19, Created)
DT 01-DEC-2001 (Tremblrel. 19, Last sequence update)
DT 01-DEC-2001 (Tremblrel. 19, Last annotation update)
DE HMUR PROTEIN.
GN HMUR.
OS Bradyrhizobium japonicum.
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Bradyrhizobium group; Bradyrhizobium.
OX NCBI_TaxID=375;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=110SPC4;
RA Nienaber A., Hennecke H., Fischer H.M.;
RT "Discovery of a haem uptake system in the soil bacterium
RT Bradyrhizobium japonicum.";
RL Submitted (MAY-2001) to the EMBL/GenBank/DBJ databases.
DR EMBL; AJ311165; CAC38746.1; -.
SQ SEQUENCE 782 AA; 83713 MW; F8FC1EBFE568AA87 CRC64;

Query Match 6.4%; Score 312.5; DB 2; Length 782;
Best Local Similarity 20.3%; Pred. No. 2.3e-09;
Matches 187; Conservative 113; Mismatches 308; Indels 311; Gaps 37;

QY 30 RAGSEAIQVLEDVHVHAKRVPKDKKVFTDARA---VSTRQDIFKSSLENLDNIVRSIPG 85
Db 71 QAQSAAPVQTLDTITVAA---TKTREIDA LAPVSSISLDQIQGLQPNFLSDVHFHVP 127
QY 86 AFTQODKSSGIVSLNIRGDSGFGGRVNTWVDGITQTFYSTSTDAGRAGSSQFGASVDSNF 145
Db 128 VSFQERGGDDPATVINIRGLQDFGRVAVVVDGARQNYORT---GHNANGSFF---LDPEL 180
QY 146 IAGLDVVKVSFSGSAGINSLAGSANLRTLGVDVVOGNNTYIGLLKGLTGNTSKGNAMA 205
```



```

Db 181 IGVVDVVRGTANTYGGAGLGLVFTKDKINDVLRGGRWVDLSGYSNNRGLGSV 240
QY 206 AIGARKWLESGAVGLYHSHRSRAQYRVGGGQHGFGAEYLERRRKQRYFVQBGGL 265
Db 241 FGGVRATPDVDIFGGAYV----RTQGNK-DGNGTEIGTNGQ-----VEAGLM 284
QY 266 KFNNSCKWDERFORPWYKTKYQKYNDOELOKYIEGHKSKWRENAP-OYDITPID-- 322
Db 285 KLT-----VRP-----ALGHEVKFGAVPDQYQYDIGQPNRG 315
QY 323 -----PSSLKQKQAGNLT-----FKLEYDGVFNKYTAQPRDLTKIGSKII 363
Db 316 PTTTOALLALNRGSSVYASDAKNYSGTVTNYSPLSPDLNFDWHSVY-----GNTDN 368
QY 364 NRYQFNYGLSLNSYAN-----LNLTAAYNSGRQYKPKSGKFTGWGLLKDFETYNNAKILD 419
Db 369 DQTKYHYGTTPSAYCNGFGNNVSGCVGDKR-----GVVL-----NTYG-----VD 410
QY 420 LNTATFRLPRETELOTLTGLFNHYFHEYGNRPPEELGLFEDGPDQDNGLYSLGRFKGD 479
Db 411 ANNTRENV-----GDMRNALTWGVDAFQDD 436
QY 480 -----KGLLPQKSTIVQPAQ-----SQYENTFYFDA--ALKKDIYRLNY--STN 519
Db 437 VIITDSRG-----NSNITTPSGIRTSGFGLQKQNYST-WLEAVSAIRYDRYDLDSGKTS 491
QY 520 TVGYRFGGTYGYSDDEPKRAFGENSPYKKHCNOSCGIYEPVLKYYKKRRANNHSVS 579
Db 492 TGGDRFSPKIT-----LGV-TPV----- 508
QY 580 ISADFGDYFEPFASRTHRWPNIOEMFYSGIDSGVHTA----- 619
Db 509 -----PGFPYYSIAGYRAPSTETVI-----SGAHATGGGPAFPVCPDGTAGLFCFL 557
QY 620 -----LKPERANTWQFGNTYKKGILLKODDTLGLKLGVYRSRIDNYIHNVYKQWMDLNGNI 675
Db 558 PNPNLREVEGKNKEVGLNLYDNIFSANDSFRGKINLFRNVDSDYI-----DLVASA 609
QY 676 PSWVSSTG-LAYTIOHNRNFKDKVHKHGFLELNYDYGRFTNLSYAY-QKSTOPTNFSDA 733
Db 610 PVAVPPGSGFSQFYQYQNIAN-ARIRGFEAETMYDAGDFIGVAGHYIQGKNVATNIGLA 668
QY 734 SESPNNAKEDQLKQVGLSRVSALPRDYGRLVGTWGLNKLTLGAMRYFKGKSIRATA 793
Db 669 TIPT-----RKVVTTGGVRLDORTLLTA 692
QY 794 EERYIDGTNGNTSNRQLGKRISIKQETILARQPLIFDFYAAVEPKKNLIFRAEVKNLFD 853
Db 693 QWASYGPNNDVPAGYLPATGYE-----LVNLYLTYNATRDIVLSASIDNLLN 739
QY 854 ---RRYIDPLDAGNDAATQ 869
Db 740 QYVRPYAIP-GSSTDGTQT 757

RESULT 15
Q9KLM6 PRELIMINARY; PRT; 784 AA.
AC Q9KLM6;
DT 01-OCT-2000 (TrEMBLrel. 15, Created)
DT 01-OCT-2000 (TrEMBLrel. 15, Last sequence update)
DT 01-DEC-2001 (TrEMBLrel. 19, Last annotation update)
DE TONB RECEPTOR-RELATED PROTEIN.
GN VCA0625.
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=EL TOR N16961 / SEROTYPE O1;
RX MEDLINE=20406833; PubMed=10952301;
RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwinn M.L.,

```

```

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,
RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,
RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,
RA McDonald L., Uterback T., Fleischmann R.D., Nierman W.C., White O.,
RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,
RA Fraser C.M.;
RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio
RT cholerae."
RL Nature 406:477-483(2000).
DR EMBL; AE004392; AAF96526.1; -.
DR TIGR; VCA0625; -.
DR InterPro; IPR000531; TonB_boxC.
DR PROSITE; PS00430; TONB_DEPENDENT_REC_1; UNKNOWN_1.
KW Receptor; Complete proteome.
SQ SEQUENCE 784 AA; 87975 MW; 605DCFF12B0CBB10 CRC64;

Query Match 6.3%; Score 308; DB 16; Length 784;
Best Local Similarity 21.2%; Pred. No. 41e-09;
Matches 202; Conservative 115; Mismatches 314; Indels 320; Gaps 46;

QY 6 RLKPCIFLMGVMLYHSHYAEADAGRAGSEAIQVLEDVHVKAKRVPKDKKVTTDARAVST 65
Db 36 QLKPL-FTLLPVVLSVVQAQE---NTEQAVDETVTVH-----GQSILTDQR---T 79
QY 66 QDIFKSSENLDNIVRSIPGAFTQODKSSGIVS-----LNIRGDSGFGRVN 111
Db 80 RSLDLK-----VRGIANA---DIFSGITSVQSNMHNHEAGALDIGIRGVGGEGRVP 127
QY 112 TWVDGITQTFVSTDAGRAGSSQFASVDSNFIAGLDVVVKGS-----FSGSAGINS 164
Db 128 IFIDGSLQ---SHTSRGYQGVSDR--TYIDTLLSSLTYNKGATIESSPVASGAVG--- 179
QY 165 LAGSANLRTGLVDVVOGNNNTYGLLLKGLTGTNSTKGNAMAAIGARKWLESGASGVLYG 224
Db 180 --GVNATTLGIKDIKDDQAFGVVLKARANHNRTPDV-----SGDYSEOGQ-----YA 227
QY 225 HSRRSVAQNVYRG---GGGQHIGNFGA--EYLERRQRYFVQEGGLKFNSNSGKWERDF 278
Db 228 LDERGHSATKHSGLMLGLYQAESFNTVLAYSKRSGKNHF-----AGK----- 271
QY 279 QRPYWKTKWYQKYNP-----QELQKVIIEGHDKSWRENLAPOYDITPIDPSSLKQQSAGN 333
Db 272 -----KGVEEQEPVVGQGVVNTSFSD-SW----- 298
QY 334 LFKLEYDGVFNKYTAQPRDLNTKIGSKRIINRYQFNYGLSLNSYANLNTAAYNSGRQK 393
Db 299 LFKLASD---TGTAHNADFNYRHQAQAGEVLMAYWYKSSDWEGN-----PYPDGKDR 349
QY 394 YPKGSKFTGWGLLKDFETYNNAKILDLNNTATFRLPRETELOTLTGLFNYPHN-----EY 447
Db 350 MPQ-----WGL-----GTAKV-----NTYSANYYPQDHPWLN 378
QY 448 GKNRFPPEELGLFDGPDQDNGLYSLGRFKGDKGLLPQKSTIVQAPAGSOYFNTFYDAA- 506
Db 379 NANFWYTEADL-----AQYGLWA-LG-----TNAEQFPAHYHNDRSG 415
QY 507 -----LKKDIYRLNYSNTVGYRFGGTYGYGSDDEKRAFGENSPYKKHCNQSC 558
Db 416 LSLTNETLLTQWPVRLNLYGLAQOQNERLSPEEDG---QTRFTKT----- 455
QY 559 GIYEPVLKYYKKRRANNHSVSISADFG-----DY----- 587
Db 456 ---VTSRHGKRTAQNLFANADIDYSPRLVQLGLNLHNAKSTDYQTKQQLDYKEKLDLL 510
QY 588 -----FMP-----FASYSRTHRMENIOEMFVS-OIGDSGVHTALKPERANTWQEGENTYK 636
Db 511 SEFTYALTPTQLFLKSRTRYRMPSLSETTILSNVFSVNPYNPKPEQAWNNEVGVOFMA 570
QY 637 KGLLKQDDTLGLKLGVYRSRIDNYIHNVYKQWDLNGLNIPSWSVSSTGLAYTIOHNRNFKDK 696
Db 571 SNSVLQDRLNLVSYPFNSIKDFISG--GRLAKTGP-MSEW--QANFTFT---NY-DK 620

```

